Random Effect Models in the Statistical Analysis of Human Fecundability Data: Application to artificial insemination with sperm from donor.

Thesis

UNRESTRICTED

Random Effect Models in the Statistical Analysis of Human Fecundability Data.

Application to artificial insemination with sperm from donor.

Submitted for a Ph.D. degree
Tuesday 11th March 1997
Discipline: Life sciences (Biometrics and Demography)

Open University, U.K.,
Sponsoring Establishment : MRC Biostatistics Unit,
Cambridge, U.K.

Ecochard René, M.D.

Date of submission: 11th March 1997
Date of award : 16th June 1997

ProQuest LLC.
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106 – 1346

Random Effect Models in the Statistical Analysis of Human Fecundability Data.

Application to artificial insemination with sperm from donor.

Ecochard René, M.D.

Submitted for a Ph.D. degree
Tuesday 11th March 1997
Discipline: Life sciences (Biometrics and Demography)

Open University, U.K.,

Sponsoring Establishment : MRC Biostatistics Unit, Cambridge, U.K.
Director of Studies : D.G. Clayton
Other supervisors : J. Esteve and A. Pickles

Examiners : H. Goldstein and D. Spiegelhalter

CONTENTS

ABSTRACT

The main aim of this dissertation is to explore methodological approaches to correlated binary data and to assess their suitability for the analysis of data on human fertility. The dataset concerns a study of Artificial Insemination by Donor (AID). AID represents an unusual research opportunity to study both male and female fecundability simultaneously. In each attempt to conceive, artificial insemination is carried out in consecutive ovulatory cycles until conception or change of treatment. The probability of conception may differ between women, so that the data are discrete time survival data with censoring and between-subject heterogeneity. There is also potential heterogeneity between donors. Non-systematic allocation of the donor to recipient ensures that the same woman receives semen from several donors. This added heterogeneity as well as other cycle dependent covariates have to be taken into account. The analysis must also take account of covariates, most of them time-varying. Our dataset have a crossed hierarchical structure due to the presence of both, female and male factors. The rather complicated "design" calls for unit specific regression models. These models are presented as well as their lack of tractability except in some rather specific cases. The motivation for choosing Gaussian random effects in unit specific regression models is discussed. We demonstrate the use of an approximate inference method (Penalized Quasi Likelihood). This method is shown to be a useful and practical way of carrying out preliminary data analysis. Finally a Bayesian procedure (Gibbs sampling) provides validation and more accurate results despite the intensive computation it needs.

The main substantive finding of the analysis is the unexpectedly pronounced heterogeneity of donor fecundability, even after inclusion of conventional measures of sperm quality into the model. These measures were shown to be predictive at the donor level but not at the level of individual donation.

INTRODUCTION

The main aim of this dissertation is to explore methodological approaches adapted to correlated binary data and to assess their suitability for the analysis of data on human fertility.

The dataset concerns a study of Artificial Insemination by Donor (AID). AID is a treatment for couples suffering from male infertility : the treatment consists of insemination with sperm from a donor, the husband being not fertile. The statistical unit will be the "cycle", also called ovulatory (or : "menstrual") cycle. An ovulatory cycle begins with menses and finishes either with menses or with a pregnancy in case of conception. In this dataset during each observed cycle an insemination takes place and thus the "cycle" can be considered as a trial resulting either in a success (if the conception occurs) or in a failure (if not). It is a binary process.

That study of artificial insemination by donor represents an unusual research opportunity to study both male and female fertility simultaneously. In "normal" couples, these aspects are nearly totally confounded, while in these data the non-systematic allocation of the donor to recipient allows the effect to be differentiated.

Women remain under observation until conception or change of treatment occurs and probability of conception may be different between women. Therefore data can be described as discrete time survival data with censoring and between subject heterogeneity. There is also a potential heterogeneity between donors in their ability to provide a "good" sperm. The non-systematic allocation of the donor to recipient allows a same woman to receive sperm from more than one donor. This added heterogeneity due to the donors as

well as other cycle dependent covariates have to be taken into account in our analysis. Introduction of time dependent covariates at the subject specific level into discrete time survival data is one of the partly unresolved issue that will be discussed in this dissertation. Our dataset have a crossed hierarchical structure due to the presence of both, female (ovulatory cycles within woman) and male factors (inseminations within donors). This dissertation will alternatively focus on two types of models for heterogeneity, one for event occurrence data concerning the recipients and the other dealing with the overdispersion of success rates among donors. But after having done separate analyses, we will have to deal with the complete crossed hierarchy : this will be a second challenging characteristic of the AID problem.

After a rapid literature review (Chapter 1), and the presentation of our dataset (Chapter 2), a marginal model for discrete time survival data in heterogeneous population (Chapter 3) will be used to describe the decline of success rates (hazards) consecutively to the withdrawal of the more fertile women after their success and relate these hazards to the observed covariates. Then a more specific description of the manifestations of heterogeneity among the women and among the donors will be done and the condition to introduce covariates into these overdispersion models will be presented (Chapter 4). The rather complicated design call for unit specific regression models. These models will be presented in Chapter 5 as well as their lack of tractability except in some rather specific cases. In the last Chapters the motivation for choosing Gaussian random effects in unit specific regression models will be discussed. We shall demonstrate the use of an approximate inference method (Penalized Quasi Likelihood approach) both, for a separate analysis of female and male hierarchies (Chapter 6), and for the analysis of the crossed hierachies — female and male — (Chapter 7). This approximate inference method will be

shown to be a useful and practical way of carrying out preliminary data analysis. Finally a Bayesian procedure (Gibbs sampling) will provide validation and more accurate results despite the intensive computation it needs (Chapter 8). The last Chapter (Chapter 9) will presents briefly some additional aspects selected for their practical implications or as potential areas for further developments.

Chapter 1 Literature review and context

In this Chapter we shall only present an initial outline of the medical background to AID and of the statistical approaches in demographic literature. Throughout the following chapters further details will be given as necessary.

1. Medical background to AID

AID is a treatment for couples suffering from male infertility (Lansac, in Gray et al, 1993). In France, a network of 20 sperm-banks called CECOS (Centre for the study and preservation of Eggs and Sperm) has been established to serve the entire country. These centres collect donations and preserve the sperm for artificial insemination by donor.

Recipients

Only couples are accepted. Moreover, all evidence for a sterility of the women would be a contra-indication and thus the women involved in our study are considered as normally fertile. An added aspect of selection is related to the husband : The main reason for using AID is the infertility of the husband : either there are no spermatozoa in the husband's sperm or just a few which are barely mobile. This aspect of the selection will be discussed further.

Allocation of donor to recipients

There is no systematic assignment of donor to recipient and sperm from several donors is used in the course of a treatment. Nevertheless, care is taken to avoid any difficulty with

blood group incompatibility and with any visible evidence against paternity from the husband — colour of the skin and of hair, etc... (Gray et al., 1993).

The treatment

The treatment consists of successive ovulatory cycles with insemination until a pregnancy is obtained, but after 12 cycles of failure another treatment is proposed. The insemination applies the following rules : the optimal day for conception is determined using the basal body temperature (BBT) curve, and the clinical examination of the cervix. When it is time to perform the insemination, the straw containing the frozen sample is removed from the liquid nitrogen and allowed to thaw. The semen is slowly injected directly into the cervix or more rarely into the uterus (intra-uterine insemination).

The outcome

The ovulatory cycle with insemination is considered to be a success if the woman conceives. Some of the resulting pregnancies are interrupted by a miscarriage. In that case, but also after a successful birth of a child, another attempt of inseminations is proposed to obtain another pregnancy : a new series of ovulatory cycles with insemination is then observed, for the same woman.

Fecundability

Fecundability was originally defined as "the probability for a *married* woman to conceive during a month, in the absence of any Malthusian or neo-Malthusian practice intended to limit procreation" (Gini, 1924). The conception was later defined as the fertilization of an ovum by a sperm (United Nations, 1958). Since the definition has led to various interpretations, Bongaarts (1975) made the following definitions :

- total fecundability (TF) is the probability that any conception occurs during an ovulatory cycle ; this includes non-implanted fertilized ova and conceptions aborted spontaneously before the end of the cycle.

- recognizable fecundability (RF) is the probability of a conception which is recognizable at the end of the conception cycle by the non-occurence of the menstruation.

- effective fecundability (EF) is the probability of a conception which will end in a live birth.

In the AID dataset the cycle of insemination is considered to be a success if the woman conceives, this diagnosis being based on 21 days of hyperthermia, or on biological or echographical evidence. The fecundability we considered here is therefore the recognizable fecundability.

Female and male "fecundability"

We should note that studies of artificial insemination by donor represent an unusual research opportunity to study both male and female fertility simultaneously. In "normal" couples, these aspects are nearly totally confounded, while in these data the non-systematic allocation of the donor to recipient allows the effect to be differentiated. For simplicity, in the following we will use the word "fecundability" for both female and male participation in the success of an ovulatory cycle with insemination.

Prognostic factors

Success rates are highest for women who are less than 35 years old, are married to azoospermic husbands and have no fertility problems (Gray et al, 1993, p 240). In the literature we have much evidence for a heterogeneity between the women (recipient) after

controlling for observed (known) covariates : the success rates are considerably lower for women inseminated for a first pregnancy than for women seeking a second or third pregnancy (CECOS Federation and Lelannou, 1987).

Our data analysis, presented below , will confirm these accepted notions.

2. Statistical approaches of the fecundability in demographic literature.

Fecundability was defined by Gini (1924) and Henry (1961) at a "subject specific" level, that is to say in relation to a specific couple, rather than a group. In this Section we will present the long lasting interest of the demographers in subject specific models but also the difficulty they encounter when they have to introduce covariates into these models.

Subject specific models

Demographers and more generally people working on natural fecundability did pioneering work in statistical models taking account of the heterogeneity between couples : Gini (1924) introduced the notion of the diversity of fecundability in human, Sheps (1964) gave a clear description of the mixtures of geometric distributions for discrete time models, and Vaupel et al (1979) denoted "frailty" the variation in the baseline hazard from subject to subject due to omitted covariates. The titles of successive books published in this field reflect the connection with other areas of statistics ; one of the last was precisely entitled "Demographic Applications of Event History Analysis" (Trussel et al, 1992).

Mixture of the geometric distribution

Gini (1924) was the first to formulate conceptive delays analytically by use of mixtures of the geometric distribution. Couples attempting pregnancy are to be followed for up to a small number of menstrual cycles, or until pregnancy occurs. Aging of the members of the couple during the follow-up will have negligible effects on its fecundability : the conception probability for each couple is taken as constant over the time.

Models involving heterogeneous populations represent a closer approach to reality than do those that assume homogeneity. This was pointed out as early as 1924 by Gini. The first effort to combine the geometric distribution with some distribution for heterogeneity was made by Tietze et al (1959) who combined the geometric distribution with an arbitrary three-point distribution of fecundabilities, which he fitted by trial-and-error. Henry (1961) suggested substituting a type I curve (beta distribution) to represent the way fecundabilities vary among couples. Sheps (1964) investigated extensively the characteristics of rates of conception in a mixture of geometric distribution. Potter and Parker (1964), Majundar and Sheps (1970), Singh and Bhaduri (1972), Maruani and Schwartz (1983), Weinberg and Gladen (1986) and others developed moment methods or likelihood methods for estimation of less or more complicated extensions of the beta mixture of geometric distribution. More recently Heckman and Walker (1990) gave a new insight into this field. They demonstrate the following points : [i] there are some patterns of declining marginal fecundability which cannot be explained by a mixture of geometric distributions; [ii] if the fecundability is constant for each couple, the population distribution of fecundability can be consistently estimated either parametrically or as a mixture of a finite number of population with a given fecundability; [iii] it is possible to test non-parametrically for the presence of percentage of totally sterile women.

Covariates

As will be shown later in this dissertation, it is rather difficult to include in a same model heterogeneity and covariates. Demographers propose various solutions to include covariates in their waiting time models. Two main directions have been considered: [i] introduction of the covariates in parametric models, extension of the beta-geometric model to study the effect of covariates on marginal conception rates (Weinberg and Gladen, 1986) — see Chapter 4 for more details — [ii] use of semi-parametric models. The latter approach has been motivated by the need to use a discrete time scale (the ovulatory cycle) and when time dependent covariates exist. In these circumstances,Weinberg et al, (1994)

> "prefer to abandon the parametric modelling approach in favour of semiparametric regression modelling of marginal hazards leading to a Cox's regression model for life tables in discrete time".

Continuous time scale

Avoiding the difficulties related to the discret time scale, some authors model the time as continuous, having then at their disposal other, parametric and semi parametric, models. For a fully parametric model they describe the heterogeneity among couples using the gamma distribution. They obtain a gamma-exponential model (e.g., Singh and Bhaduri , 1972) ; Sheps and Menken, 1973; Boldsen and Schaunberg, 1990). In this approach, although the time scale is genuinely discrete, it is assumed that conception takes place in continuous time, i.e., that a woman can conceive at any time, not only once in each menstrual cycle.

Semi parametric models for delay until conception on continuous time scale are used when the delay to conception is calculated over long periods of time. They are closely related to

frailty models for survival data and allow the covariates to change with time (change of "treatment status" during the course of the study) (e.g., Larsen and Vaupel, 1993).

None of these solutions are fully adequate in the present context where, the delay until conception being short, there is little justification for a continuous time scale. Very recently demographers have published some results obtained using a random effects model for cycle viability in fertility studies on discrete time scale, (Zhou et al., 1996). This is an application of the recent development of the random effect models in the literature.

3. Random effects models

The two aspects of heterogeneity in the AID dataset call for two different bodies of statistical literature : Whereas the woman heterogeneity calls for (discrete time) survival data and "frailty" models, the principal end-point being the occurrence of conception, the donor heterogeneity calls for "overdispersion" models for binary data, the number of cycles per donor being considered as independent of the success rate of the donor. In our context it will be useful, for a better understanding, to generate these models by a "causal" hierarchical model.

The literature on survival models with frailty parameters has a long history (see reviews by Clayton, 1988; Andersen et al., 1992; Pickles, 1994). The need to introduce an extra *random component* into survival data was first formulated in the field of continuous bivariate survival time data (Clayton, 1978 ; Vaupel et al, 1979). This random component was introduced as a *random multiplier* on the hazard scale and termed "*frailty*". This will be discussed in detail in Chapter 5.

In the description of the overdispersion models for binary data the roles of fixed and random effects has been a challenging aspect. Williams (1982) pointed out the desirability of placing the "heterogeneity" as additive on the same scale as the fixed effect : this

condition maintains the "unit-treatment additivity" (Cox, 1984). This will be discussed in Chapter 4. The specification of the distribution of overdispersion is still also an active area of research (Lee and Nelder, 1996). Nevertheless, analytical tractability and the flexibility to model real data are conflicting requirements (Clayton, discussion of Lee and Nelder, 1996) :

> "conjugate distributions are rarely available and, even when they are, additivity of fixed and random effects on the same scale is often not possible if tractability is to be maintained"

The Normal distribution has a rather particular place in this area as distribution of the random factor(s). Normal random effect distributions allow one to model complex inter-dependency between units. This will be discussed further in this dissertation (see Chapter 6 and following).

Linear and non-linear hierarchical models (Goldstein, 1991) are currently an extremely active area of research in biostatistics : they provide a useful frame to model complex data such as AID data. Hierarchical models appear in the literature under a variety of titles. In sociological research, they are often refered to as multilevel linear models (e.g., Goldstein, 1986, Mason, 1983). In biometric applications, the term mixed-effects models or random-effects model are common (e.g., Laird and Ware, 1982). In the statistical literature, they are often referred to as variance components models (e.g., Searle et al, 1992). As regression models for hierarchical data (or e.g., multiple measurements from several individuals) they are often presented as random intercept models.

A further step is to think of ways in which statistical techniques should take more completely the hierarchical structure into account including all connections and interactions between levels (Goldtein, 1986 ; Rasbash and Goldstein, 1994 ; Bryk and Raudenbush, 1992). These aspects will be discussed in Chapter 6 and 7.

Chapter 2 The dataset

The dataset concerns the 1901 women treated from January 1985 to March 1994 at the Center for Study and Preservation of Eggs and Sperm (CECOS Department of Biological Reproduction, HCL, France, Pr J.C. CZYBA and Dr D. COTTINET), with Artificial Insemination with Frozen Donor semen (AID). We will first present the hierarchical structure of the dataset and the pregnancies. Then a general description of the characteristics of the women, donors and cycles will be made. Finally an analysis of the censoring process will be performed in order to study the extent to which it could distort the inference.

1. Hierarchical structure of the dataset

Female hierarchy

A total of 12100 cycles was observed in 2437 "attempts" by 1901 women. At each attempt, a woman is inseminated at each of a consecutive series of ovulatory cycles, until success or right censoring. No attempt is prolonged past 12 cycles. Successful women ask often (one third) for a new attempt. Moreover, despite unsuccessful attempts a part of the women ask for a new attempt some months or years later. Figure 1 A presents this hierachy.

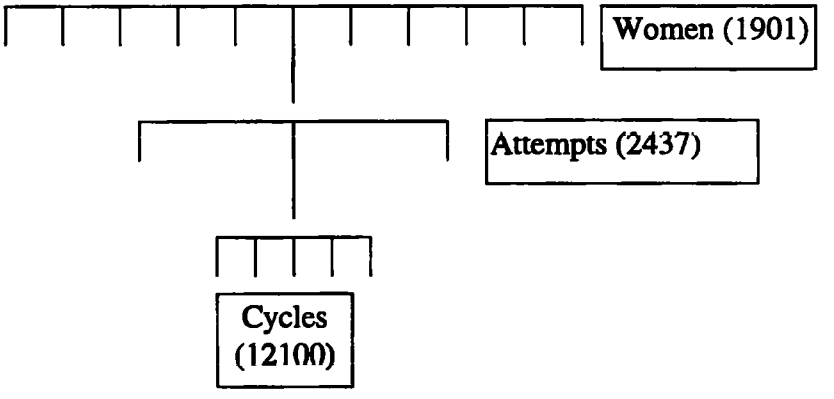


Figure 1 A Female hierarchy

Selection of the women

We have to deal with a very strong process of selection. Couples are "triply selected" (Leridon, 1984) : First, couples are selected by the time they have been waiting for a conception without any success. A second selection is made by the physician, who has decided which couples are good candidates for this specific treatment. The third selection is the self-selection determined by the couples themselves when they decide to see a doctor, and accept or refuse the treatment.

More specifically, [i] the sterility proportion among the women is a priori lower than in the general population because the infertility of the husband is primarily responsible for the sterility in AID, [ii] two different groups of woman can be identified : women whose husbands are sterile (no spermatozoa in sperm) and women whose husbands are hypofertile (just a few spermatozoa which are barely fertile). These two categories of women have different levels of fecundability. It was shown that women are globally less fertile if their husband are not sterile, among the patients recruited for AID.

In our data, about two thirds (1178/1901) of the husbands suffer from male sterility, while one third of husbands have just a few spermatozoa which are barely mobile. It will be important to introduced this distinction in the model specification. Moreover, the women whose husbands are not totally sterile could be more homogeneous having been selected as less fertile, and thus the residual heterogeneity would differ between groups. This will be investigated in Chapter 7.

Male hierarchy

Sperm donations were made by 279 different donors. These donors have made a total of 1 328 donations, collected over about a month's time from a single donor. Each donation is divided into a number of aliquots (about 20-30) which are separately frozen as straws which will be used for each insemination.

Figure 1 B shows the male hierarchy, and the Table 1 the number of donations per donor.

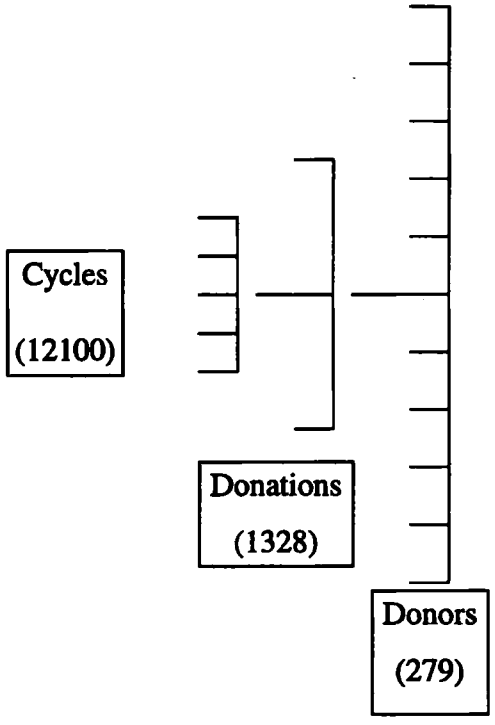


Figure 1 B Male hierarchy

Selection of donors and donations

CECOS policy states that semen donation is an anonymous gift for which no payment is received. Only married men under the age of 55, who have one or more normal children and whose spouse consents, are accepted as donors. A selection is necessary to eliminate

three types of donors. Firstly, infertile and subfertile donors are rejected : fatherhood may be regarded as a guarantee of the donor's fertility, the ultimate criterion, however, is semen analysis. Secondly, donors carrying an infectious disease which might present a risk to the recipient and/or the child are not accepted. Thirdly, men carrying a hereditary disease which might present a risk to any child that is conceived are also not accepted as donor: this can be easily eliminated by genetically screening donors, both with karyotype and family and personal histories. Donations of poor quality on initial testing are discarded and not used for insemination.

The number of inseminations per donor may be influenced by the knowledge the physician have of his success rate. The physician can decide to stop further use of sperm of a donor after a series of unsuccessful inseminations with semen of that donor. Chapter 4 will discuss this selection process and illustrate the resulting selection bias. Chapter 7 will show how a unit-specific model may help to correct for it.

Donations	Donors	(percentage)
1	21	(7.5)
2	23	(8.2)
3	32	(11.5)
4	47	(16.8)
5	59	(21.1)
6	44	(15.8)
7	29	(10.4)
8	12	(4.3)
9	6	(2.2)
10 - 13	6	(2.2)

Table 1 Number of donations per donor.

In about 97% of the cases less than 50 straws were frozen (in liquid nitrogen) per donation. Table 2 shows the number of straws per donations.

Number of straws	Donations	(percentage)
2 - 9	96	(7.2)
10 - 19	453	(34.2)
20 - 29	460	(34.7)
30 - 39	214	(16.2)
40 - 49	61	(4.6)
50 - 59	19	(1.4)
60 - 69	12	(0.9)
70 - 79	6	(0.5)
80 - 83	4	(0.3)
unknown	3	

Table 2 Number of straws per donation.

In the majority of the cycles (84%), two inseminations were carried out. In 96.4% of these cases the second was timed in the 48 hours following the first one. And in every cases a same donation was used for both inseminations of a same cycle.

Crossed hierarchy

There is no systematic assignment of donor to recipients but some association as the result of the calendar time. We give a global overview of this through figures 1 to 3. Cycles are organized into two overlapping hierarchies as showed in Figure 1 C.

Figure 2 and Figure 3 show the relationship between donors and women, each point representing one insemination.

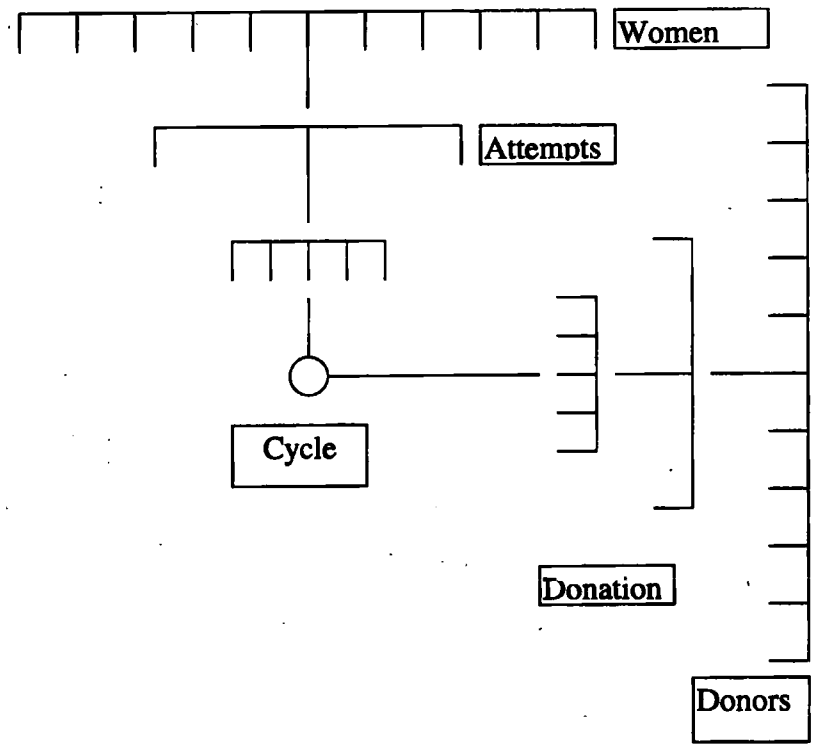


Figure 1 C The crossed hierarchical structure

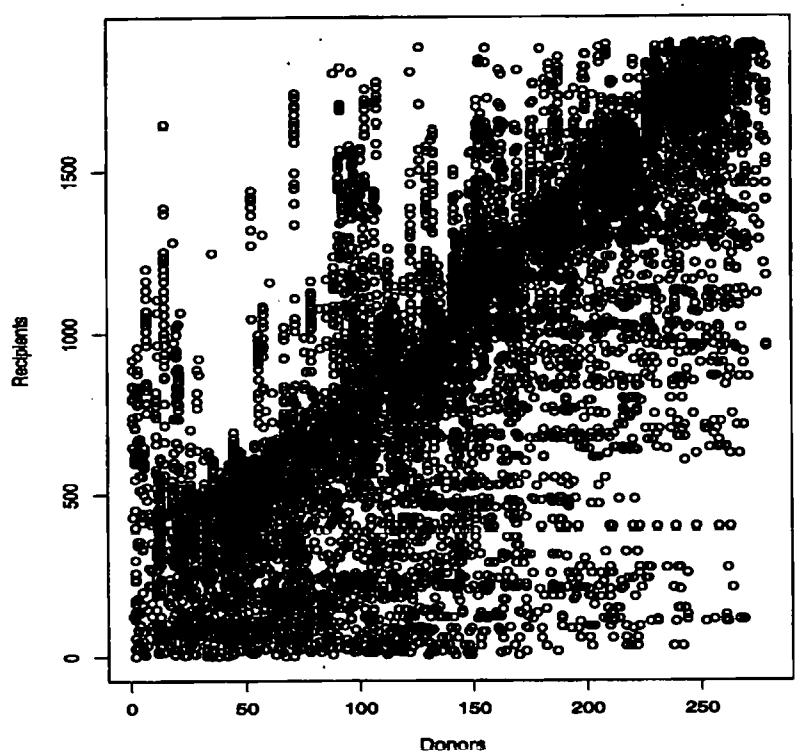


Figure 2 Calendar time relationship on the "diagonal" : donor and recipient are contemporary. Donor and recipient numbers are their calendar order of registration.

In Figures 2 and 3 the lower right part of the figure concerns recipients asking for more than one attempt or beginning late after their registration (the identification number is attributed at the date of the first visit). The upper left corner points can be read as allocation of "old sperm" to women registered later.

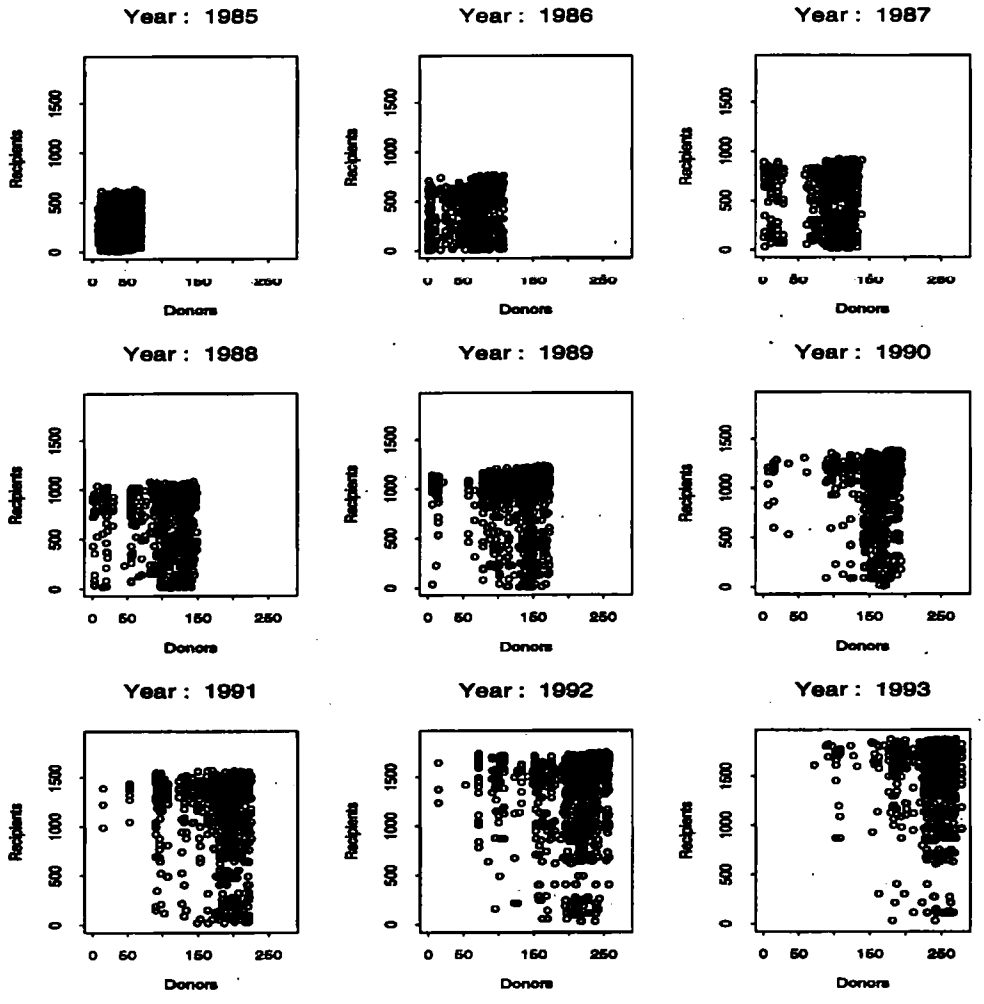


Figure 3 Donor versus recipient rank by year of insemination

There is a lag of about 5 years between donations (1980-1990) and inseminations (1985-1993). But there is a temporal relationship between ranks of recipients and donors.

Table 3 shows the number of cycles per "pair" of woman-donor.

Cycles per "pair"	woman-donor "pairs"
1	6 347
2	1 850
3	480
4	111
5	21
6	7
7	2
8	1

Table 3 Distribution of number of insemination cycles with the same woman-donor pairing. Complete dataset.

Sperm from each donor was used for insemination of several women. However there is a small degree of association : the situations where the same woman receives sperm from a same donor was quite frequent. Intuitively, this coincidence is rather larger than might be expected under random allocation, analysing this association in detail would be a major computation exercice. The association is probably due to the fact that the number of donors being available at a given time is rather low. Additionally the necessity to avoid any difficulty with blood group and with visible evidence against paternity (colour of the skin, of hair, etc..) limits the choice slightly.

2. Pregnancies

909 / 1901 (47.8 %) women obtained a pregnancy withing the first series of ovulatory cycles -the first attempt. A total number of 1213 pregnancies was obtained when

considering all attempts, either ending with the birth of a child or interrupted by a miscarriage. Table 4 shows the distribution of these various outcomes.

Outcome	Pregnancies
Normal	898
Miscarriage	146
Extra-uterine Pregnancy	9
Pathological events during pregnancy	7
Stillbirths	5
Elective termination	1
Unknown	147

Table 4 Outcome of the 1 213 pregnancies.

The outcome of the pregnancy is known for 1 066 of the 1 213 pregnancies. The missing data are partly due to current pregnancies and partly due to loss to follow-up after the first trimester.

There is a suggestion that miscarriage or still-birth is more likely when the conception occurs in later cycle. During the first attempt, if the conception occurred at the first cycles only 8.5% of the pregnancies ended with a miscarriage, but if the conception occurs after the first cycle 15 % of the pregnancies result in miscarriage. However, most pregnancies have a favourable outcome and the main reason for repeated attempts is to achieve a larger family size. Table 5 provides for each of the successive attempts, the total number of ovulatory cycles — with insemination —, the number of successes — pregnancies — and the selection of women asking for a further attempt.

Attempt	Cycles	Women	Successes	Failures	Further attempt among successful	Further attempt among unsuccessful
1	9740	1901	909	992	362	70
2	1902	432	244	188	74	16
3	378	90	54	36	11	2
4	69	13	6	7	1	0
5	11	1	0	1	-	-

Table 5 Attempts, successes and failures. Complete dataset.

84% (362/432) of the women beginning a second series have been successful during the first attempt and as such have demonstrated their fecundability. These women, after a first success (conception), ask for another series of inseminations either because the conception was followed by a miscarriage (93, i.e., 26.7 %) or to have another child (255, i.e., 73.3 %). Table 6 shows a pronounced decrease in conception probability over time. It is not properly viewed as a time effect within woman, but as a selection effect in a heterogeneous population of women. The success rates appear to be higher during the second attempt. This progressive selection of the women under study will be studied at length in Chapter 4.

	First attempt		Second attempt	
Cycle number	Successes/cycles	Success Rates	Successes/cycles	Success Rates
1	194/1538	0.1261	72/432	0.1667
2	136/1332	0.1021	40/337	0.1187
3	119/1176	0.1012	36/279	0.1290
4	101/1022	0.0988	28/223	0.1256
5	76/914	0.0832	19/172	0.1105
6	66/806	0.0819	13/138	0.0942
7	49/660	0.0742	10/102	0.0980
8	58/584	0.0993	13/85	0.1529
9	32/503	0.0636	10/61	0.1639
10	33/454	0.0727	2/37	0.0541
11	25/402	0.0622	1/24	0.0417
12	20/349	0.0573	0/12	0.0000

Table 6 Conception rates over time. First and second attempt.

3. Covariates

The data having been recorded in a prospective way, most of the known prognostic factors have been carefully registered. These covariates characterize the women, their cycles, but also the donors and their donations.

Women and cycles

The 1901 women were registered between 1985 and 1994, as shown in Table 7.

Year	Women
1985	478
1986	198
1987	184
1988	166
1989	170
1990	159
1991	185
1992	184
1993	144
1994 (3 months)	33

Table 7 AID. Complete dataset. 1901 women. Year of first insemination.

For a number of reasons, including new treatments after 1985 (In Vitro Fertilization with sperm of donors or not) and lack of donors in the recent period, the recruitment of patients has decreased since 1985. Nevertheless this higher number of "first inseminations" in 1985 is mainly a first sign of the left truncated nature of the dataset; this phenomenon will be discussed later.

Table 8 shows the age of the women when they began the treatment, with the dramatic decrease after age 38.

Age (woman)	Women
18 - 24	125
25 - 29	775
30 - 34	732
35 - 38*	246
39 - 43	23

Table 8 Age of the women at their first insemination.

* Generally, after age 38, insemination is not proposed to infertile couples, the probability of success decreasing and the risk of foetal abnormality being higher at that age.

These women are normally fertile. As shown by Table 9, about 97% of the women had evidence of ovulation before treatment. For the other 3% the basal body temperatures curves (BBT) was not relevant and the decision was made to consider the women as fertile.

BBT before treatment	Women	(percentage)
Correct	1 472	(77.4)
Irregular cycles	365	(19.2)
Anovulatory	40	(2.1)
Non interpretable	24	(1.3)

Table 9 Basal body temperature (BBT) before treatment.

An overall judgement of the quality of the cycle was made at its end by the physician, based on the basal body temperature curve (BBT). This classification is presented in Table 10. Nevertheless, in most of the cases these data were recorded by a physician who knew the outcome and thus must be interpreted with some care.

BBT curve	Cycles	(percentage)
Normal	11 197	92.5
Abnormal	718	6.0
Non-interpretable	185	1.5

Table 10 Characterization a posteriori of the BBT curve.

As shown in Table 11, about 2/3 of inseminations took place on the ovulation day, the day before or the day after. The "ovulation day" is the estimated day of ovulation according to the basal body temperature and hormonal levels. It must be recalled that the length of the

ovulatory cycles varies from cycle to cycle and between women and that this is the reason why the insemination cannot always be optimally timed.

Insemination day First insemination	Insemination	(percentage)
- 5 or less	475	(3.9)
- 4	391	(3.2)
- 3	825	(6.8)
- 2	1 585	(13.1)
- 1	2 499	(20.7)
0	3 853	(31.8)
+ 1	1 631	(13.5)
+ 2 or later	841	(7.0)

Table 11 Day of first insemination relatively to the "day of ovulation".

In 38.6% of inseminations the sperm was inserted directly into the uterus, either routinely — some of the physicians prefer this procedure — or after one or more ovulatory cycles exhibiting a mucus of poor quality index.

Situation of insemination	Inseminations	(percent)
Cervix	7 428	(61.4)
Intra-uterine	4 672	(38.6)

Table 12 Situation of the insemination.

The Insler score is partly related to the quality of the mucus, partly to the opening and aspect of the cervix. This score increases when the observation takes place closer to the ovulation, but also characterizes the ability of the cervix to produce a good mucus. Table 13 shows that in about 85 % of the cycles the Insler score was acceptable (7 or more).

Insler score	Inseminations	(percent)
5 or less	650	(8.8)
6	541	(7.3)
7	1 130	(15.2)
8	1 943	(26.2)
9	3 164	(42.6)

Table 13 Insler score. Cervical inseminations (n = 7 428).

Donor and donations

As shown in Table 14, a majority of donors are between 30 and 39 years old (72 %)

Age (donor)	Donors	(percentage)
26 - 29	22	(7.9)
30 - 34	97	(34.8)
35 - 39	104	(37.3)
40 - 44	43	(15.4)
45 - 53	13	(4.7)

Table 14 Age of the donors

The "quality" of the donations is described by the number and mobility of the spermatozoa and also the "post-thaw index". Despite strong correlations between these three parameters they give independent information of the quality of the sperm and have long been considered to be reliable indicators.

Number of spermatozoa 10^6/ml	Donations	(percentage)
7 - 49	151	(11.4)
50 - 99	532	(40.1)
100 - 149	351	(26.4)
150 - 199	164	(12.3)
200 - 249	62	(4.7)
250 - 299	30	(2.3)
300 - 349	24	(1.8)
350 - 399	6	(0.5)
400 - 449	6	(0.5)
450 - 500	2	(0.2)

Table 15 Number of spermatozoa per aliquot.

Percentage of mobile spermatozoa	Donations	(percentage)
30	21	(1.6)
40	315	(23.7)
50	473	(35.6)
60	355	(26.7)
70	126	(9.5)
80	38	(2.9)

Table 16 Mobility of the spermatozoa.

All donations had a good percentage of mobility; as previously stated others were discarded and not used for insemination for evident reasons. The post thaw quality index is set on an ordinal scale and describes the number of mobile spermatozoa under the microscope after thaw (this verification being made once per donation).

Post thaw quality index	Donations	(percentage)
1	2	(0.2)
2	41	(3.1)
3	67	(5.0)
4	286	(21.5)
5	136	(10.2)
6	472	(35.5)
7	102	(7.7)
8	172	(13.0)
9	50	(3.8)

Table 17 Post thaw quality index

4. Censoring and left trunction

Our data are discrete time survival data : women are recruited at their first ovulatory cycle with insemination and remain under observation until conception occurs. But some women

were midway through an attempt when data collection commenced, in 1985. Moreover, some women were withdrawn from the cohort before occurrence of any conception.

Late entry - left truncation

There is some left truncation. The few left truncated cases are a consequence of the inclusion of women having been treated for a few months at the beginning of the study (1985).

These late entries do not create any selection "bias" if [i] they are registered as such with the correct position on the time scale — cycle rank —, [ii] there is no added process of selection, i.e., if they have the same probability of conception as the women they join.

The entry time (cycle) is shown in Table 18.

Entry to study (cycle rank)	Women	(percentage)
1	1 538	(80.9)
2	132	(6.9)
3	36	(1.9)
4	43	(2.3)
5	25	(1.3)
6	22	(1.2)
7	22	(1.2)
8	25	(1.3)
9	22	(1.2)
10	9	(0.5)
11	15	(0.8)
12	12	(0.6)

Table 18 Entry times First attempt.

Right Censoring

In AID -discrete time scale- the censoring occurs between two cycles : the subsequent cycles will not be included in the analysis, but the result — conception or not — at the end of the previous cycle is known.

There is some right censoring, in a few cases due to an "interval pregnancy". Table 19 shows the reasons for end of observation (first attempt).

Cycle	Pregnancy	Interval pregnancy	Other treatment (IVF)	Adoptions	Medical reason*	Decision to stop**	End point***	NA
1	194	1	3	1	4	1	13	14
2	136	1	5	2	5	2	11	12
3	119	1	15	0	5	3	16	9
4	101	2	3	0	4	5	10	14
5	76	3	7	0	5	5	7	16
6	66	1	34	3	13	13	9	8
7	49	1	10	1	7	10	6	7
8	58	2	11	2	8	5	5	8
9	32	0	9	2	4	5	7	5
10	33	1	4	1	4	9	5	18
11	25	0	4	4	4	8	7	23
12	20	2	4	2	1	315	0	0
TOTAL	909	15	109	18	64	381	150	134

Table 19 Reasons for end of observation (First attempt).

* Medical reasons are psychological (21) or physical (43).

** Other decisions to stop were taken in common, physician + patient (323) or by the couple (58).

*** Are classified as censored consecutively to the end point the couples without insemination for less than two years and without record of the decision to stop the treatment.

This censoring arises either because of the end of the study (end point), or because no more inseminations took place or were registered in the dataset. The first case — end point — can be compared with a type I censoring (Kalbfleish and Prentice, 1980), where the censoring time of each individual is fixed in advance. This censoring does not interfere with the inference process, and thus can be ignored. The latter — no more inseminations take place — depends arbitrarily during the course of the study on previous conceptions, and on values of observed covariates : if bad prognosis factors are observed and no success occur during the first cycles, not rarely a change of treatment is proposed before the twelfth cycle, because of physician advice or of the impatience of the couple or their lassitude. It is for example due to the decision to withdraw from the cohort the women being more advanced in age, after a few cycles, and to propose to them another treatment. This censoring process changes the distribution of the characteristics of the women under observation during the further cycles without disturbing the relation between these characteristics and the outcome. This selection process is ignorable.

A "selection bias" would be created if withdrawal process were related to future events!

> "non-response is non-ignorable if it depends on an unobserved response" (Little and Rubin, 1987).

As we shall show below, interval pregnancies create such a selection !

Interval pregnancies

Sometimes, after one attempt and before commencing a subsequent one, a woman may conceive naturally. These quotes "interval pregnancies" are quite rare (15 interval pregnancies are registered). These are causes of censoring, not competitive events during an attempt. It is highly probable that the higher the probability to conceive under insemination the greater is also the chance to be withdrawn because of an interval

pregnancy. These women are thus right censored because of indirect information about there fecundability and thus formally create a selection bias. This selection process is non-ignorable. Nevertheless, the quite low rate of these pregnancies allow us to ignore this phenomenon for practical purposes.

At the end of this chapter it appears that the AID dataset has potentially a very complex correlation structure resulting from repeated attempts to conceive by the same woman, and repeated use of sperm donors. After a conception occurs the woman is removed from observation at least for a few months : conception acts as a "selection process" (Leridon and Spira, 1984). The proportion of less fertile couples increases as the waiting time lengthens.

Moreover important explanatory variables change during the course of the study : Insler score, day of insemination and pharmacological stimulation (Clomiphen citrate) are all time-varying explanatory variables. Besides if bad prognosis factors are observed and no success occurred during the first cycles, quite often a change of treatment is proposed before the twelfth cycle and the woman is censored.

Nevertheless we have seen that, except for the very specific — and rare — case of interval pregnancy, the strong selection process at the origin of the dataset is compatible with the use of conditional inference.

Chapter 3 Marginal models for discrete time survival data

Our data are censored discrete time survival data : women remain under observation until conception or censoring occurs. For simplicity in this Chapter we ignore male heterogeneity and the multivariate nature of the data and consider only the first attempt at pregnancy by each woman.

The models we consider in this Chapter are marginal models. These models are used to relate the marginal hazards to observed covariates. We point out two important facts. First, we show that, despite the introduction of observed covariates, a decrease of the baseline remains which may be interpreted as a sign of the presence of other, unobserved, prognostic factors. Second, the fact that the parameters of this marginal model have no interpretation at the subject level is stressed.

1. The logistic model for marginal rates

Let us define the discrete "hazard" at cycle t, λ_t, as the probability of conception at cycle t, (successful insemination) conditional upon the fact that no conception occurred in previous cycles.

$$\lambda_t = \Pr(T = t | T > t-1)$$

We choose to treat the time scale as discrete, because the menstrual cycle is a logical, physiological unit of time for the phenomenon under study. The time measurement represents the number of attempts required to conceive, an ordinal scale.

Let us write λ_t as a function of prognosis factors in the following way,

$$\log\frac{\lambda_t}{1-\lambda_t} = \beta_{0t} + x_t^T\beta_1 .$$

where, on a logit scale β_{0t} is an arbitrary location parameter corresponding to cycle t and β_1 are the vector of changes in the fraction of a positive response for a change of one unit in the corresponding covariates.

The marginal rates are modelled as a function of covariates without explicitly accounting for subject heterogeneity. The regression coefficients have interpretation for the population rather than for any individual and hence we will use the term "population-averaged" (PA) model in this case (Zeger et al,1988).

2. Likelihood

The construction of the likelihood for censored failure time data on a discrete time scale will be progressive. Our final objective is to cover complex situations, including time dependent covariates and unobserved heterogeneity. This potential complexity calls for a unified method of presentation, with the drawback of the need to use a rather more technical notation than for a simpler situation where a simpler explanation would be sufficient. We will successively deal with the likelihood, formed as the product of conditional contributions, and the factorization of the likelihood in case of censoring and the existence of two equivalent ways to write the likelihood of censored data. Chapter 4 will add the modifications needed to cover the case of unobserved heterogeneity.

Following Gelfand and Smith (1990), distributions are denoted by brackets, so joint, conditional, and marginal forms, appear as $[X,Y]$,$[X|Y]$, and $[X]$. Multiplication of

densities is denoted as a product $[X,Y]=[X|Y][Y]$. The process of marginalization is denoted by integration of the conditional distributinon such as $[X]=\int[X|Y][Y]dY$.

2.1 The likelihood, formed as the product of conditional contributions

The response (delay until conception) considered as polytomous, with 13 categories

Suppose that the response of each woman in the study is the delay until conception, T say, or the absence of conception during the first 12 months. Initially let us ignore the cases where some censoring takes place before the 12*th* cycle.

In this case the response can be considered as polytomous, with 13 categories, T=1,...12,13, the 13*th* for women having not conceived during the 12 cycles of insemination. Time of conception or censoring is thus integer valued within 1-13.

Let Y_t be an indictor of $T=t$; Y_t is the outcome of the tth insemination for one woman. Finally $Y_{(t)}$ will be used as a shorthand for $(Y_1, Y_2, ..., Y_t)$: $Y_{(13)}$ carries the same information as T.

Product of cycle contributions

The likelihood of these polytomous data is expressed as a product of cycle contributions. Let $[Y_1, Y_2, ..., Y_t]$ be the related distribution of the time ordered random variable. Following rules of conditional probability, the multinomial distribution can be expressed as a product of factors, the conditional probabilities to conceive during each cycle, the conception having not taken place earlier.

$$[Y_{(13)}]=[Y_1][Y_2|Y_{(1)}][Y_3|Y_{(2)}]\cdots[Y_{13}|Y_{(12)}]$$

Thus the likelihood is formed as the product of conditional contributions, of the whole study group, over successive cycles. The likelihood can now be written as

$$L = \prod_{1}^{13} [Y_t | Y_{(t-1)}]$$

2.2 Factorization of the likelihood

Let us introduce now the censoring during the first *12* cycles. C_t representing censoring at the end of a cycle *t*, i.e., between two cycles (the woman was inseminated at cycle *t* but not at cycle *t+1* and after), as before

$C_{(t)}$ will be used as a shorthand for $(C_1, C_2, ..., C_t)$

Let *H* represent the complete history of conceptions, covariates, and censoring

$$H_{(t)}=(Y_{(t)}, X_{(t)}, C_{(t)})$$

The likelihood can be constructed as a product of the conditional terms.

$$[Y_t, C_t | H_{(t-1)}, X_t]$$

The likelihood can now be written as

$$L = \prod_{1}^{12} [Y_t, C_t | H_{(t-1)}, X_t]$$

Note that following simple rules of probabilities

$$[Y_t, C_t | H_{(t-1)}, X_t] = [Y_t | H_{(t-1)}, X_t][C_t | H_{(t-1)}, Y_t, X_t]$$

and thus the full likelihood factorizes into two parts (event process, censoring process) :

$$L = \prod_{1}^{12} [Y_t | H_{(t-1)}, X_t] \prod_{1}^{12} [C_t | H_{(t-1)}, Y_t, X_t]$$

Independence

Let us study the first factor on the right side :

$$\prod_{1}^{12}[Y_t \,|H_{(t-1)}, X_t]$$

If $\left[Y_t \middle| H_{(t-1)}, X_t\right] = \left[Y_t \middle| Y_{(t-1)}, X_{(t)}\right]$ the censoring is said to be independent, because censoring does not select women in regard to their specific probability of pregnancy. Under this assumption the distribution of Y_t given the covariates is the same among censored and uncensored individuals; we can therefore drop $C_{(t-1)}$ from $H_{(t-1)}$.

Unobserved heterogeneity could be a reason for non-independent censoring ! As stated in Chapter 2, a "selection bias" would be created if the withdrawal process both within and between different attempts selects women of higher or lower fecundability. The interval pregnancies are important to consider in this context: the censoring resulting from the withdrawal of women after an interval pregnancy excludes some highly fertile women, and thus, is a case of non-independent censoring. Nevertheless interval pregnancies are not numerous and we will ignore this fact in the rest of this dissertation.

Non-informative censoring

Let us study the second part of the likelihood equation, corresponding to the censoring process $\prod_{1}^{12}[C_t|H_{(t-1)}, Y_t, X_t]$

If this contribution does not depend on the parameter of interest, the censoring is non-informative. In the AID dataset the censoring depends arbitrarily on the prognosis factors observed during the previous ovulatory cycles : a woman having bad Insler scores, for example, can be excluded from AID cohort for that reason. The relations between these

prognosis factors and the censoring process are not related to those existing between them and the outcome, and thus there is no informative censoring in our dataset.

Factorization

The likelihood inference results solely from ratios of the likelihood function for various values of the parameter (Edwards, 1972). Under the assumption of independent censoring, inference may be based only on the observed part and remain consistent when the likelihood is limited to the first factor and fully efficient if the censoring process is non-informative.

Note that even in the case of informative censoring there will be an associated loss in efficiency but no bias.

2.3 Two equivalent ways to write the likelihood of censored data

We should note that two equivalent ways can be used to write the likelihood of our censored data.

First, as arising from one unit per woman : the likelihood contribution for a woman observed until cycle t is given by the marginal distribution function for uncensored observations or the marginal survivor function for right-censored observations :

$$f_t^{\ \delta}\ F_t^{1-\delta}$$

where $\delta = 1$ if conception, and $\delta = 0$ for censoring and f_t and F_t are respectively $Pr\ (T = t)$ and Pr $(T>t)$.

Second, the likelihood may be writen as arising from one unit per cycle, the likelihood being formed as the product of conditional contributions, as stated in the previous Section. The likelihood contribution for a woman i (i=1,...,n) observed on cycle t is

$$\lambda_{it}^{\delta_{it}}(1-\lambda_{it})^{(1-\delta_{it})}$$

Since $F_t = \prod_{u \le t}(1-\lambda_u)$, these likelihood are equivalent and will be used interchangeably.

This will have a renewed interest in subject-specific models, all cycles of a same woman sharing a common basal risk.

3. Logistic analysis of the data :

In this section we present the fit of the logistic model presented above. We include all observed covariates.

Note :

SAS PROC LOGISTIC was used to fit the model :

```
proc logistic;
model y= cycle2 cycle3 cycle4 cycle5 cycle6 cycle7 cycle8
cycle9 cycle10 cycle11 cycle12
inscent  numcent mobcent agewcent
J_3orles  jplus1or clomiphen azoo deccent  mobdec numdec;
where attempt=1;
run;
```

Metric covariates are standardized (it is for example the case for the covariate azoospermia : "azoo") to have unit standard deviation

This analysis concerns only the first series ("attempt") of cycles of the women.

Table 20 shows the results.

Parameter	Estimate	SE
Intercept	-2.016	0.079
Woman level		
Age of the woman	-0.104	0.036
Azoospermia of the husband	0.090	0.036
Cycle level; female covariates		
Insler Score	0.236	0.043
Early insemination(before Ovulation day minus 2)	-0.139	0.042
Late insemination(after Ovulation day)	-0.109	0.037
Stimulation of ovulation with clomiphene citrate	-0.095	0.038
Cycle level; male covariates		
Number of spermatozoa in the semen before freezing	0.135	0.033
Percent of mobile sperm. in the semen before freezing	0.182	0.037
Global Index of quality of the semen after thaw	0.120	0.039
Cycle rank		
2	-0.253	0.120
3	-0.269	0.125
4	-0.285	0.132
5	-0.438	0.144
6	-0.486	0.151
7	-0.582	0.169
8	-0.318	0.160
9	-0.760	0.200
10	-0.628	0.198
11	-0.832	0.223
12	-0.945	0.245
- 2 log likelihood	5800.1	

Table 20 First attempt. Logistic regression. With all observed covariates

The discussion of the effect of observed covariates is delayed until Chapter 7.

As previously said, the pronounced decrease of the base-line over time (-2.53,...,-0.945) is interpretable as a selection process in a heterogeneous population. This decrease of baseline despite the introduction of observed covariates, indicates the presence of other, unobserved, characteristics. Our population of women remains clearly heterogeneous despite the introduction of the observed covariates.

4. Link function

For marginal models the choice of the link is mainly a matter of quality of fit and of interpretation of the parameters (log-odds for the logit link). In this dissertation the choice of the link function has a specific interest due to the fact that we will have to introduce both fixed and random effects, and we will look for mixing distributions with optimal properties. For this reason, we have to describe in more details the link functions which will be considered.

The following regression models relate cycle - and woman - "hazard" (which is a success in fecundability data), λ_{it} (i for woman and t for cycle) to possibly time dependent covariates x_{it}

$$\log\left(-\log\left(1-\lambda_{it}\right)\right)=\beta_{0t}+x_{it}^{T}\beta$$

for the complementary log-log or "proportional hazards" model

and

$$\log\frac{\lambda_{it}}{1-\lambda_{it}}=\beta_{0t}+x_{it}^{T}\beta$$

for the logistic regression model

and

$$\log\lambda_{it}=\beta_{0t}+x_{it}^{T}\beta$$

for the log linear regression model

the last one having been proposed by Weinberg et al (1994) , as providing for estimation of a "fecundability ratio" which is the ratio of the cycle-specific conception probability for

exposed women, divided by that for unexposed women. This last model will not be described later on. We prefer to focus on the logistic and the complementary log-log, both being used in the following Chapter after some modifications to include random effects.

The complementary log-log model

The probability of failure varies as an exponential function among individuals :

$$1-\lambda_{it}=(1-\lambda_{0t})^{e^{(x_{it}^T\beta)}}$$

This model can be obtained specifying that

$$-\log(1-\lambda_{it})=-\log(1-\lambda_{0t})e^{x_{it}^T\beta}$$

i.e., to place the covariates as acting multiplicatively on the $-\log(1-\lambda)$ scale.

$-\log(1-\lambda_{0t})$ is positive and thus can be written $e^{\beta_{0t}}$. It becomes

$$-\log(1-\lambda_{it})=e^{\beta_{0t}+x_{it}^T\beta}$$

or, which is equivalent

$$1-\lambda_{it}=\exp\left(-\exp\left(\beta_{0t}+x_{it}^T\beta\right)\right)$$

or,

$$\log\left(-\log(1-\lambda_{it})\right)=\beta_{0t}+x_{it}^T\beta$$

this last form being at the origin of the name "complementary log-log model".

If either the base-line hazard or the covariates are fixed over time the model simplifies :

for fixed base line

$$F_{t_i}=(1-\lambda_0)^{\sum_{u=1}^{t_i}\exp x_{iu}^T\beta}$$

for fixed covariates

$$F_{t_i} = F_{0t_i}^{\exp x_i^T \beta}.$$

These formulae have the same aspect as the corresponding ones for the proportional hazard model for continuous time scale. This complementary log-log model will be considered again below (Chapter 5) with the introduction of a random multiplier or "frailty" to represent the specific risk of each woman.

The logistic regression model

An alternative discrete model specifying a linear log odds model for the hazard probability at each potential conception time has been used in the previous Section. The hazard for woman *i* at cycle *t* is given by

$$\frac{\lambda_{it}}{1-\lambda_{it}} = \frac{\lambda_{0t}}{1-\lambda_{0t}} e^{x_i^T \beta}$$

Note that this model is multiplicative in $\frac{\lambda_{0t}}{1-\lambda_{0t}}$, which, being positive, can be written

$e^{\beta_{0t}}$

It becomes

$$\frac{\lambda_{it}}{1-\lambda_{it}} = e^{\beta_{0t} + x_i^T \beta}$$

or, which is equivalent

$$\log \frac{\lambda_{it}}{1-\lambda_{it}} = \beta_{0t} + x_{it}^T \beta$$

This is a linear logistic model with an arbitrary logistic location parameter corresponding to each cycle.

Comparison

The base-lines, λ_{0t}, $-\log(1-\lambda_{0t})$ and $\frac{\lambda_{0t}}{1-\lambda_{0t}}$ are very similar for *small* values of λ_{0t} as shown in Figure 4. These similarities for low hazard rates are particularely interesting in our situation : pregnancy rates in AID data set are at about 10 p cent and thus these models are very similar in our context. Nevertheless, they have a different interest : it will be shown below that the complementary log-log model gives rise to some important simplification in calculations in our discrete time scale situation.

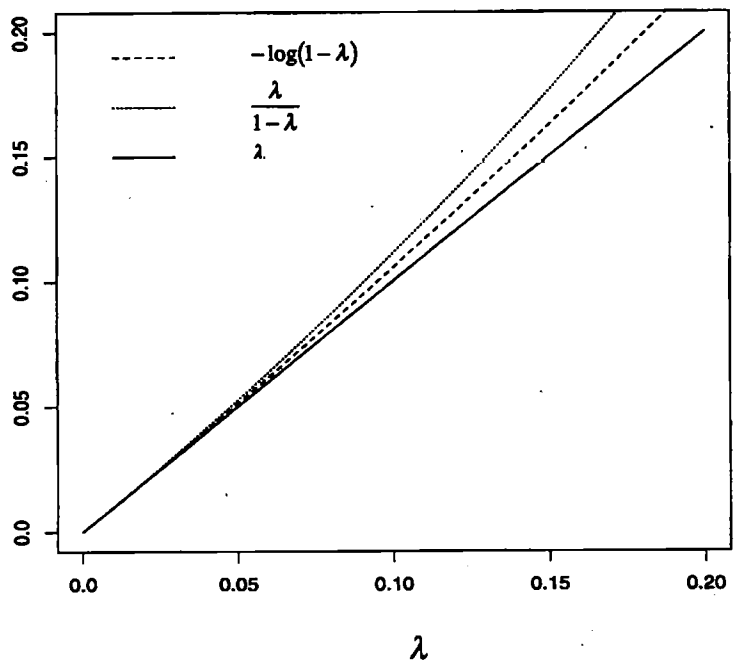


Figure 4 A graphical comparison of the three link functions

5. Interpretation of parameters of the Population Averaged model

It is wise to distinguish the size of the effect of the covariates on each subject's probability to conceive and the size of its effect on the "population" : in the presence of heterogeneity, these two effects are not equivalent (Zeger et al, 1988).

* "Subject-Specific" (SS) parameters, β_{SS} say, describe the effect of the covariate on the probability of conception of each woman,
* "Population Averaged" (PA) parameters, β_{PA} say, describe the effect of the covariate in the global result when the two groups of women, having and not having the characteristic, are compared.

In the previous Section we have estimated β_{PA}. We have verified the absence of interaction between time and each of the marginal effects. Now, a progressive selection of the less fertile women arise from the first to the twelfth cycle; therefore the absence of interaction means that the effect of the covariates does not vary with the underlying fecundability and thus our marginal results can be generalized.

The question of the interpretation of our estimates at the woman level arises naturally in our context : with regard to a specific woman what is the effect of the covariates? In the presence of heterogeneity, β_{SS} are higher than β_{PA} : the random effects variability shrinks the subject specific fixed effects parameters toward 0. The progressive selection of less fertile women modifies the relation between these two quantities! As a consequence the relationship between β_{SS} and β_{PA} is not stable, except for a very specific distribution of the heterogeneity — positive stable distribution (Hougaard, 1986) — for which the marginal decrease of fecundability does not come along with decrease of its variance among the women.

Thus, coefficients obtained using marginal models are interpretable as predictors of marginal rates but not at the subject level. Perhaps it should be pointed out that these coefficients might not be applicable in a different population with more or less heterogeneity.

At the end of this chapter it may be said that the need to take account of the heterogeneity is clear. For the most part we shall use random effects models rather than marginal models for two reasons. Firstly, the correlation in the data must be taken into account when estimating standard errors of parameter estimates. In this application the correlation structure is too complicated to be taken into account using jackknife, bootstrap or Huber's formula, as a result of the presence of correlation not only within blocks but also between blocks (donor heterogeneity). Secondly, the magnitude of "unexplained" variance components is of at least as much interest as the covariate effects, most of which being already well understood. Finally we will prefer to use a "causal" hierarchical model — or "subject" specific model — in which the unobserved heterogeneity is explicitly specified and thus take directly account of the correlation structure of the data.

Chapter 4 From the observation of overdispersion to the specification of a mixed model

Historically, the motivation for random effects models for binary data has been the observation that the variability among binary responses may exceed the variability which would be expected due to binomial variability alone. This overdispersion may be produced by correlation between binary responses or by cluster of observations with similar probability of success.

The idea of heterogeneity among women was the most likely explanation for the observation of a marginal decrease of the pregnancy rates (Gini, 1924). We shall present a further aspect of heterogeneity among women which is detected through the association between two waiting times to conception in successive attempts of the same women. Because of heterogeneity two processes corresponding to successive attempts of the same women "sharing a common basal risk" (Clayton, 1978) will resemble each other more than two other processes.

Donors being selected as fertile, the literature hardly discusses possible heterogeneity among them. In our dataset we observe however two manifestations of heterogeneity among donors : overdispersion of counts of successful fertilization by donors, and, rather unexpectedly, a marginal increase of success rates for donors whose semen has been previously used for a higher number of insemination.

In this chapter we will describe these manifestations of heterogeneity in our data set and then consider a way to specify models which takes them into account. For simplicity, we

will study, heterogeneity between women ignoring male heterogeneity and conversely, heterogeneity between donors ignoring female heterogeneity.

Note that the proportion of successful cycles obtained per woman in one attempt is not of the same nature as the proportion of successful cycles obtained per donor. For women, the numerator — number of pregnancy during one attempt, 0 or 1 if success—, and the denominator — number of cycles — are both random and related. Under the assumption that a woman has a constant conception probability, the distribution of the delay until conception is *geometric* -a special case of the negative binomial distribution. If T is the delay until conception and λ the risk (probability of success) for each cycle

$$\Pr(T = t) = \lambda(1-\lambda)^{t-1}$$

Under the assumption of homogeneous conception probability (among women and among donors) the number r of successes for a given donor is *binomial*. If R is the number of successes, m the number of ovulatory cycles for which the semen of this donor has been used, and λ the risk (probability of success) for each trial

$$r \sim Binomial(m, \lambda)$$

$$\Pr(R = r) = \binom{m}{r}\lambda^r(1-\lambda)^{m-r}$$

Despite this apparent difference between these two points of view, presenting the observation by woman or by donor will lead to equivalent inferences if inference is based on the likelihood. If t is the number of cycles observed and $\delta = 1$ for conception and $\delta = 0$ for censoring, the likelihood contribution for each woman is

$$\lambda^{\delta}(1-\lambda)^{t-\delta}$$

This likelihood takes the "binomial" form despite the fact that the model is not a binomial one.

As regards to the number of successes per donor, if the the process of conception and the censoring process -decision to stop using semen of a donor for further insemination- are independent and non-informative[1] , the likelihood contribution for each donor is binomial.

$$\lambda^r(1-\lambda)^{m-r}$$

Thus the likelihoods have the same binomial "form" despite the fact that the model are different (Cox and Hinkley, 1974, p 40).

1. Manifestations of heterogeneity among the women

We shall present the first manifestation of heterogeneity, the marginal decrease of hazards, rather briefly since it was discussed in the previous Chapter. We shall then discuss in detail a second consequence of heterogeneity among women which is the association between waiting times to conception in successive attempts.

1.1 Marginal decrease of hazards

The observation of the marginal success rates (See Table 6) provides two results. Firstly, the sharp decrease of the hazards during the first attempt : the marginal conception rate falls from 13.9 % for the first cycle, to 5.7% on the twelfth (Mantel-Haenszel χ^2 trend

[1] Note : it will be shown later than m is informative concerning the process of conception in our data set. Let us first ignore this fact.

test =35.66; p<0.001). Secondly, the women attempting for a second series have better results, respectively 16.7%, 9.4% and 0% (n = 12); this second trend is also significant: adjusted on the cycle rank the pregnancy rates are higher at the second attempt (Mantel-Haenszel χ^2 trend test =16.13; p<0.001).

Under the assumption of stability of the fecundability over a short period of time, these two observations are strong arguments for the existence of an heterogeneity of the fecundability between the women : subjects with high fertility are likely to conceive earlier ; this in turn will result in removal from observation ; and thus the distribution of the fecundability in those remaining under study will be modified with a decreasing mean. On the contrary the mean fecundabity among the women asking for a second attempt is higher than the overall fecundabity rate observed on the first attempt, a high percentage of them being of proven fertility.

The first and the second observation are not of the same nature. The first selection process is directly related to the conception process, the woman being systematically withdrawn from the first attempt after a conception. The second, is not so systematic. 84% of the women beginning a second series of cycles had conceived during the first one.

1.2 Association

A convenient way of studying the association follows from continuous bivariate survival time data (Clayton, 1978, Oakes, 1989).

Denoting the delay until conception in one attempt by S and the delay until conception in a subsequent attempt by T, we can measure association by the array of odds ratios.

$$\theta_{st} = \frac{\Pr(S = s, T = t)\Pr(S > s, T > t)}{\Pr(S = s, T > t)\Pr(S > s, T = t)}$$

θ_{st} may be consistently estimated by the sample odds ratios in the 2 x 2 table formed by cross-classifying all women observed at cycle *s* in the first attempt and at cycle *t* in the second attempt by whether or not they conceived at each of those cycles. Of course, there are many such statistics and there will be few data for estimation of most of them. If *s* and *t* can take values from 1 to 12, each pair of attempts constitutes 144 different tables. Empirical estimates of a constant θ may be obtained using the method proposed by Clayton (1978).

Figure 5 exhibits a grid with in each cell pregnancies and censoring for the first two attempts, among women having had at least two attempts : this will be useful to calculate estimates of θ.

Cell *s,t* tabulates women whose observation ceased at cycle *s* of the first attempt and *t* of the second. The cell contains the 2*2 table filled by tabulating the reason for stopping (conception or censoring) at each attempt. It is wise to note that in case of left truncation - for the first attempt- we have nevertheless included the cycles from the beginning of the attempt.

Figure 6 presents a grid with each cell having to be read as the component of θ_{st} : estimators of Pr(S=s,T=t) (upper left), Pr(S>s,T>t) (lower right), Pr(S=s,T>t) (upper right) and Pr(S>s,T=t) (lower left).

Figure 7 shows the estimated θ_{st} when they are available. Following Oakes (1989) let us show how to obtain estimates of θ_{st} using the elements contained in Figure 5. Each cell of the Table corresponds to a possible pair of values (s;t) for (S,T) and displays the number of women according to their behavior at the particular pair of cycles. The upper left ($n_{11}(s,t)$), upper right ($n_{10}(s,t)$), lower left ($n_{01}(s,t)$), and lower right ($n_{00}(s,t)$) entrie in each cell count

are, respectively, the double successes, censoring in T and success in S, Success in T and censoring in S, and double censoring at (s,t).

To see how θ may be estimated, define

$$
\begin{aligned}
r_{11}(s,t) &= n_{11}(s,t) \\
r_{01}(s,t) &= n_{01}(s,t) + \sum_{u>s}\{n_{11}(u,t) + n_{01}(u,t)\} \\
r_{10}(s,t) &= n_{10}(s,t) + \sum_{v>t}\{n_{11}(s,v) + n_{10}(s,v)\} \\
r_{00}(s,t) &= n_{00}(s,t) + \sum_{u>s}\{n_{01}(u,t) + n_{00}(u,t)\} \\
&+ \sum_{v>t}\{n_{10}(s,v) + n_{00}(s,v)\} \\
&+ \sum_{u>s}\sum_{v>t}\{n_{11}(u,v) + n_{01}(u,v) + n_{10}(u,v) + n_{00}(u,v)\}
\end{aligned}
$$

and $\hat{\theta}(s,t) = r_{11}(s,t)r_{00}(s,t) / r_{10}(s,t)r_{01}(s,t)$

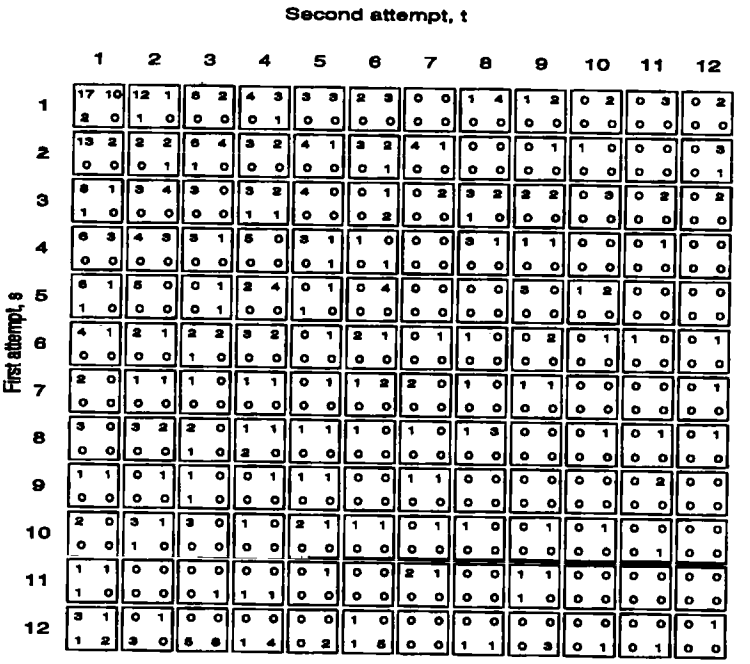


Figure 5 Pregnancies and censoring for the first two attempts, among women having undergone at least two attempts

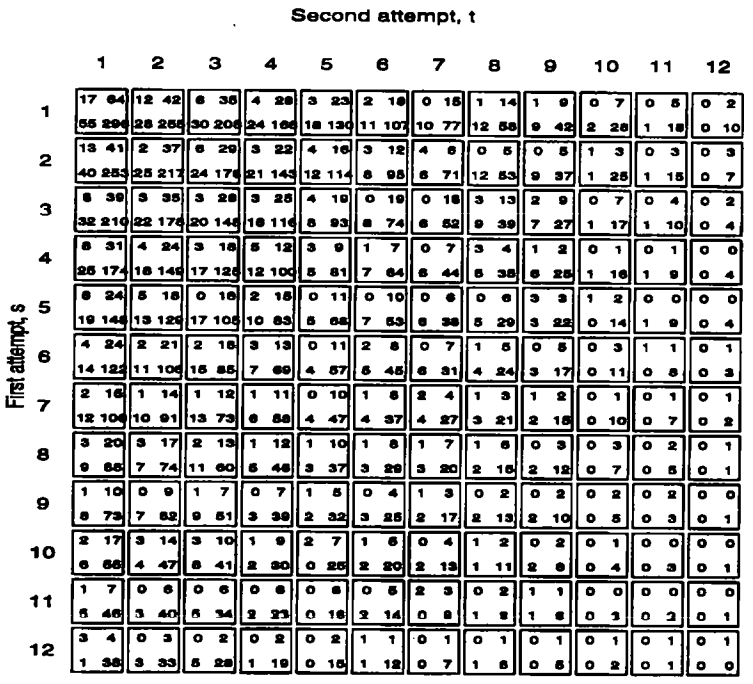


Figure 6 Two by two table prepared for calculating θ_{st}

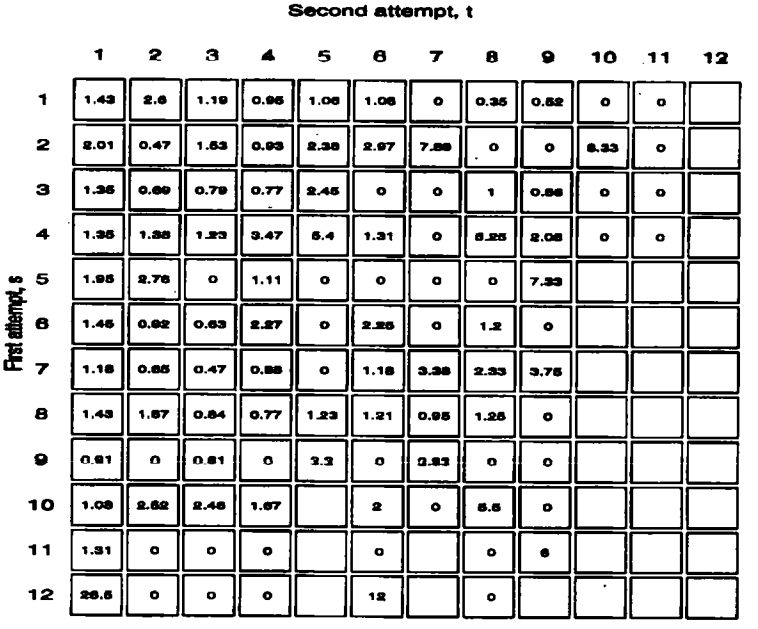


Figure 7 θ_{st} *when they are available (i.e. non-zero denominator)*

The data from these tables (Figure 7) may be combined to give a single estimate of the common odds ratio using the Mantel-Haenszel (MH) procedure. In our case, between first and second attempt

$$\hat{\theta}_{MH} = 1.73$$

To test for heterogeneity the conventional Mantel-Haenszel procedure for one degree of freedom provides us with a solution, testing against the null hypothesis ($\theta = 1$) :

$\chi^2_{1df} = 56.7$

Despite the fact that one woman could contribute to all 144 tables this test remains valid, the contributions of each table being uncorrelated under the null hypothesis.

This quasi-independence of tables cannot be assumed when the null hypothesis does not hold and thus MH confidence intervals are not correct. As a conclusion, the association provides us with a second piece of evidence for heterogeneity among the women.

The same procedure was used to estimate the association between the first and the third attempts — $\hat{\theta}_{MH} = 3.6$ — and between the second and the third — $\hat{\theta}_{MH} = 1.6$. These measures of association was estimated respectively over 432 women — between first and second — and over 90 women — between first and third, and second and third.

In the next Chapter, an interpretation of the size of the association under the gamma-geometric model will be proposed.

2. Manifestations of heterogeneity among the donors

In this Section we shall describe the overdispersion among donor success rates and the marginal increase of success rates for donors whose semen has been previously used for a higher number of insemination.

2.1 Overdispersion

Under the assumption of homogeneity (among donors and among women) , the mean and variance of the number of successes under the binomial model are respectively

$$E(r) = m\lambda$$

$$Var\,(r) = m\lambda(1-\lambda)$$

The maximum likelihood estimator of λ is the simple overall mean

$$\hat{\lambda} = \frac{\sum_{v=1}^{d} r_k}{\sum_{v=1}^{d} m_k}$$

where r_k and m_k, are respectively the number of successes and inseminations for the kth of d , (k = 1,...,d) donors. The *deviance* with respect to a model in which each donor has its own probability of success is

$$D = -2\sum_{k=1}^{d} \log \frac{\left(\frac{r_k}{m_k}\right)^{r_k}\left(1-\frac{r_k}{m_k}\right)^{m_k-r_k}}{\hat{\lambda}^{r_k}\left(1-\hat{\lambda}\right)^{m_k-r_k}} = 771, \quad \text{df=278}$$

The result may be compared with the upper bound of the χ^2 95 % confidence interval, 318. This provide a goodness of fit test of the assumption of homogeneity, which is rejected. Some care must be taken in assuming deviance to be χ^2 in this circumstance, since there is little information per donor (McCullagh and Nelder, 1989). Nervertheless the large difference between observed deviance and the upper bound provides an argument in favor of heterogeneity between donors. Heterogeneity among donors may be in part explained by heterogeneity among women. More precisely overdispersion among donors is expected from heterogeneity among women and heterogeneity may result from a non-random association between women and donors. We will see later that a part (but only a part) of this overdispersion can be explained by the observed covariates. The residual overdispersion will be considered as a consequence of the existence of other, unobserved, characteristics.

Rewriting the model to take account of this heterogeneity, we leave the mean unaffected and inflate the variance :

$$E(r) = \mu$$

$$Var(r) = \sigma^2 . m\mu(1-\mu)$$

= dispersion parameter * binomial variance

where σ^2 is independent of μ. The dispersion parameter is the so-called scale factor of GLM models. This is a first model (very simple and perhaps not a good one) for overdispersion. The variance is now the product of the binomial variance, and a dispersion parameter σ^2.

By equating the Pearson X^2 to the expectation of a χ^2 distribution having the same number of degrees of freedom, we obtain a moment estimation of σ^2:

$$\tilde{\sigma}^2 = \frac{1}{d-1}\sum_{k=1}^{d}\frac{(r_k - m_k\hat{\mu})^2}{m_k\hat{\mu}(1-\hat{\mu})} = 2.54$$

X^2 being only asymptotically χ^2_{d-1}, the estimator ($\tilde{\sigma}^2$, presented above) has a slight bias of order $O(m^{-1})$ (McCullagh and Nelder 1989 p127). $\tilde{\sigma}^2$ gives an idea of the size of the variance of the number of successes per donor. Since it is more than twice the binomial variance this is good evidence of heterogeneity among donors.

A priori, donors are withdrawn from observation when there are no more aliquot : the donations are attributed to women while the stock lasts ; as a consequence, we do not expect any trend in marginal rates. Observation of the data set will partly change our mind about this last aspect.

2.2 Marginal increase of success rates

In the previous section, the rank of the cycles among the clusters (donors) was ignored. Let us use this information. It provides another point of view to look at the heterogeneity. A phenomenon of selection -less strong than for the women, nevertheless- is observed when the successive cycles are ordered in the sequential order of the use of the sperm donations of a same donor. Table 21 shows the reality of this phenomenon.

Rank of the Insemination (per donor)	Number of Inseminations	Successes	% successful
0 - 9	2 445	242	9.9
10 - 19	2 508	228	9.1
20 - 29	2 207	213	9.6
30 - 39	1 821	169	9.3
40 - 49	1 356	130	9.6
50 - 59	883	109	12.3
60 - 69	535	75	14
70 - 79	217	32	14.7
80 -	128	15	11.7

Table 21 Risk of pregnancy according to the sequential order of the sperm of each donor.

The observed trend is significant : Mantel-Haenszel χ^2 trend test =10.58; p<0.001. Several hypotheses can be proposed to explain this apparent "improvement of the quality of the sperm".

Relation between the number of aliquots and the fecundability

As a first reason for the relation between the "rank" of insemination and the success rate could be a relation between the *number* of aliquots obtained and the quality of the sperm : a donor whose donations can produce more aliquots could also provide better results.
This assumption is partly supported by the data, as shown Table 22 (Mantel-Haenszel trend test χ^2 =9; p ≈ 0.03).

Calendar period effect

Sperm having being used for a higher number of inseminations could be sperm of the first period of donation, the more recent sperm donations having being used until now for a lower number of cycles. The better results with sperm having been used often could be due to an overall higher fertility of donors in the first period of donation (1978 - 1985 say) : this observation would be in favor of the current idea of a recent decrease of fertility. As can be seen from Table 23 this assumption is not validated by the data (Mantel-Haenszel χ^2 trend test ≈ 0.;p ≈ 0.956).

Number of aliquots	Number of Cycles	successes	% successful
0 - 9	469	42	9
10 - 19	3 082	260	8.4
20 - 29	4 212	413	9.8
30 - 39	2 609	322	12.3
40 - 49	984	101	10.3
50 - 59	298	28	9.4
60 - 69	226	20	8.8
70 - 79	143	15	10.5
80 - 89	77	12	15.6

Table 22 Success of cycles according to the number of aliquots of each donation. Complete data.

Year of recruitment of the donor	Number of cycles	Successes	% successful
79	14	3	21.4
80	55	2	3.6
82*	274	20	7.3
83	481	48	10.
84	1 934	164	8.5
85	1 172	109	9.3
86	2 854	358	12.5
87	1 308	150	11.5
88	897	65	7.2
89	1 258	124	9.9
90	1 800	166	9.2
91	33	3	9.0
92	20	1	5.0

Table 23 Success of cycles according to the year of recruitment of the donor. Complete data set.

* no donor was recruited in 1981

Selection bias

The physicians suggested to us at least two other potential explanations, both able to create a selection bias- increase of the success rate with "rank" of insemination :

First, clinicians observing more successes with some *donors* give them the preference for further use! The heterogeneous nature of the group of donors justifies their method : the posterior probability of conception is higher for donors having had successful sperm.

Second, the deliberate choice to allocate "old sperm" to potentially highly fertile *women*. For example a woman asking for a second child after a rapid success, will receive semen from a donor being no more currently "in use" : the physician expecting a rapid success for this woman give her sperm, considered as "old", considering that whichever semen is given the insemination will be successfull.

In the first case the selection is based on observation of the donor, in the second of the woman. In the second, it is noteworthy that the heterogeneity between women could create a difficulty when analysing donor effects. If later straws of some donors are donated to highly fertile women the good result could be attributed by mistake to the donor. In both case the marginal — PA — analysis is missleading. Chapter 6 will discuss this again and show how a subject specific analysis will correct the bias !

Finally we could propose a fifth hypothesis, the progressive improvement of the quality of the sperm of each donor when it remain frozen in liquid nitrogen. Some French biologists are fond of this assumption, having the opportunity to observe the same time improvement effect on wine... but our data does not support this hypothesis.

At the end of this Section we have a clear evidence for both female and male heterogeneity. In the next two Sections we will successively present the specification of a model for overdispersed binary data and then a mixed model, these Sections being a first attempt to introduce fixed and random effect in a same model.

3. Specification of a model for overdispersed binary data.

Two simple models aimed at the interpretation of the above described overdispersion will be presented : one describing the delay until conception — a mixture of geometric — , the other the number of successes among donors — a mixture of binomials.

We shall discuss first the consequence of this heterogeneity in describing it with the first two moments of the mixing distribution and in a second Section we shall see that specifying the mixing distribution can provide further insight into the problem of overdispersion.

3.1 Analysis of overdispersion without specifying the mixing distribution

Following Sheps (1964), T being the delay until conception, λ the probability of conception of a specific woman at each cycle, we suppose that λ varies from woman to woman, with $E(\lambda)=\mu$, and $Var(\lambda)=\theta$.

$$\Pr(T=1) = E(\lambda) = \mu$$

$$\Pr(T=2) = E(\lambda(1-\lambda)) = E(\lambda) - E(\lambda^2)$$

$$= \mu - \mu^2 - \theta$$

$$= \mu(1-\mu) - \theta$$

which shows that the probability of success at the second trial is less than it would be if λ were constant and equal to μ. Hence, a smaller proportion of the total group may be expected to conceive in the second month than would be the case for a homogeneous

population and the decrease of hazards is directly related to the variance of the distribution of fecundability among the women. The probability of success at the second trial given a failure at the first may be written

$$\frac{\Pr(T=2)}{1-\Pr(T=1)} = \frac{\mu-\mu^2-\theta}{1-\mu} = \mu - \frac{\theta}{1-\mu}$$

From the observation of the success rate in the first and second trial m_1 and m_2 we can therefore estimate the first two moments of the distribution of the woman fecundability :

$$\tilde{\mu} = m_1$$

$$\tilde{\theta} = (m_1 - m_2)(1-m_1)$$

The estimation of the first two moments of the distribution of heterogeneity between the women are respectively $\tilde{\mu}$ =0.13 and $\tilde{\theta}$ =0.02 (standard deviation, $\sqrt{\tilde{\theta}} = 0.145$, high compared with the mean). Note that these estimations rely only on the data obtained on these first two cycles. Estimation of mean and variance of fecundability at successive cycles may be obtained using the same methods and are shown Table 24 which suggest that both mean and variance are decreasing. Although the estimations are rather crude the decrease of the variance between the first and second cycle seems to be very sharp. This observation may be a consequence of the high level of selection arising at the first cycle, interpreted as being a consequence of the presence of a subgroup of women having a high probability of conception. Later on, the distribution of heterogeneity among women is probably smoother and unimodal. Note that the heterogeneity among women at the *t*th insemination could be unchanging with *t* or not: the stability is obtained only for some specific distributions of this heterogeneity (Hougaard, 1986 b).

Cycle *(t)*	Number of inseminations	$\tilde{\mu}_t$	$\tilde{\theta}$
1	1 538	0.13	0.02
2	1 332	0.10	0.00
3	1 176	0.10	0.00
4	1 022	0.10	0.01
5	914	0.08	0.00
6	806	0.08	0.01
7	660	0.07	-0.02
8	584	0.10	0.03
9	503	0.06	-0.01
10	454	0.07	0.01
11	402	0.06	0.00
12	349	0.06	-

Table 24 Estimation of the mean and variance of woman fecundability from two successive cycles (first attempt).

Sheps (1964) proposed to estimate higher moments of the distribution, using the information conveyed by all the observed cycles. The distribution of marginal hazards can be written entirely in terms of the moments of the distribution of λ in the initial population of inseminated women. Thus, for our AID data, this distribution can be reconstructed from its first 12 moments as soon as the sample size is large enough.

Overdispersion of the number of successes per donor

We shall discuss a two level hierarchical approach to analyse overdispersion of the number of successes per donor : at level I, $\Pr(R = r|\lambda) \sim Binomial(m, \lambda)$ and, at level II a distribution for λ among the donors, defined only through its mean μ and variance var(λ). The marginal mean and variance of r, the number of successes are respectively

$$E(r) = E\left[E(r|\lambda)\right] = m\mu$$

$$\begin{aligned} Var(r) &= E\left[\text{var}(r|\lambda)\right] + \text{var}\left[E(r|\lambda)\right] \\ &= E\left[m\lambda(1-\lambda)\right] + \text{var}\left[E(r|\lambda)\right] \\ &= mE(\lambda) - mE(\lambda^2) + m^2\,\text{var}(\lambda) \\ &= m\mu(1-\mu) + m(m-1)\,\text{var}(\lambda) \end{aligned}$$

i.e. as the sum of two components one comming from the inter-individual variance and the other from intra-individual variance

Writing var(λ) as a function of the mean, $\text{var}(\lambda) = \phi\mu(1-\mu)$ with ϕ indepedent of μ we can see that the overdispersion may be written $\sigma^2 = 1 + (m-1)\phi$

Williams (1982) proposed to estimate ϕ through an iterated weighted least square algorithm, where the matrix of weights is diag$[1/(1+(m_k-1)\phi)]$ and m_k is the number of inseminations of the *k*th of *d* donors. ϕ is estimated by equating Pearson's X^2 to its expectation, writing

$$X^2 = \sum_{k=1}^{d} \frac{(r_k - m_k\hat{\mu})^2}{\left[m_k\hat{\mu}(1-\hat{\mu})\{1+\phi(m_k-1)\}\right]}.$$

More technically speaking, this iterative algorithm proposed by Williams (1982), can be characterized by the following scheme :

1 - Estimate μ by a preliminary estimator $\tilde{\mu}^{(P)}$, assuming $\phi = 0$

2 - Estimate ϕ by equating X^2 to its expectation

3 - Using the weights from step 2, $\tilde{w}_k = \frac{1}{\tilde{\sigma}_k^2} = 1/(1+(m_k-1)\tilde{\phi})$, reestimate μ by a weighted least square using these new weights. Treating the resulting estimation as a new preliminary estimator return to step 2. Iterate until convergence.

Estimate of μ and $Var(\lambda)$ applying Williams' method to donor heterogeneity are respectively $\tilde{\mu}$ =0.096 and $\tilde{Var}(\lambda)$=0.003 (standard deviation : 0.057, quite high compared with the mean, but smaller than the same parameter concerning the women), Residual deviance =773.

This procedure provides a quasi-likelihood estimation of μ (Wedderburn, 1974; McCullagh and Nelder, 1983 and 1989) and an estimation of the variance var(λ) which may be related to the so-called pseudo-likelihood approach.

Note : Pseudo-likelihood

Pseudo-likelihood -PL- (Carroll and Ruppert, 1988; Davidian and Giltinan, 1995) : Without a probabilistic model ML can not be used to estimate ϕ. Therefore estimating equations for ϕ are obtained by considering the residuals between the observed, *y*, and fitted, $\hat{\mu}$, as normally distributed with zero mean and a variance which is defined by the relation between the variance and the mean (variance function). Pseudo-likelihood method corresponds to maximizing the normal log-likelihood evaluated at $\hat{\mu}$, which lends the procedure its name; however, as is the case for the quasi-likelihood method for the estimation of μ, PL may be regarded as an omnibus method for estimating variance components. Equating Pearson's X^2 with the number of degrees of freedom is a moment method based on the same approximation: residuals are supposed to have a Gaussian distribution and the variance parameter is obtained via a generalized weighted least-square algorithm.

3.2 Explicitely definition of the mixing distribution

We shall now use a fully specified mixing distribution. The probability of success for a given woman or a given donor will be distributed according to the beta distribution. This

full specification of the model allows likelihood estimations of the mean but also of the variance of the fecundability among women and among donors.

The beta-distribution takes on many shapes : the probability density fonction can be strictly increasing, strictly decreasing, U-shaped or unimodal (if ν and τ are both greater than 1).

$$f(\lambda)=\frac{1}{B(\nu,\tau)}\lambda^{\nu-1}(1-\lambda)^{\tau-1},\nu>0,\ \tau>0$$

$$\text{with } E(\lambda)=\mu=\frac{\nu}{\nu+\tau} \text{ and, } Var(\lambda)=\mu(1-\mu)\frac{1}{\nu+\tau+1}$$

The beta distribution is conjugate to the geometric distribution and to the binomial distribution.

If $\text{beta}(\nu,\tau)$ is a prior for λ among women, having observed t -1 unsuccessful cycles and then one success for a woman *i*, since the likelihood is proportional to $\lambda(1-\lambda)^{t-1}$ the posterior distribution $f(\lambda_i)$ is

$$\text{beta}(\nu+1,\tau+(t-1)).$$

Concerning the donors, if $\text{beta}(\nu,\tau)$ is a prior for λ, having observed *r* successful cycles over *m* insemination with a sperm donor *j*, since the likelihood is proportional to $\lambda^r(1-\lambda)^{m-r}$, the posterior distribution $f(\lambda_j)$ is

$$\text{beta}(\nu+r,\tau+m-r).$$

The mixtures are respectively the beta-geometric and the beta-binomial. Both will be presented now and applied to our data. This analysis will provide new estimations of the variance of the fecundability among women and among donors.

The beta-geometric distribution.

T being the delay until conception, the marginal probability function, $Pr\,(T = t)$ is

$$\Pr(T=t) = \int_0^1 f(\lambda)\lambda(1-\lambda)^{t-1}\,\partial\lambda$$

$$= \int_0^1 \frac{1}{B(\nu,\tau)}\lambda^{\nu-1}(1-\lambda)^{\tau-1}\lambda(1-\lambda)^{t-1}\,\partial\lambda$$

which takes a closed form named the ***beta-geometric distribution*** (e.g. Weinberg and Gladen, 1986)

$$\Pr(T=t) = \frac{B(\nu+1,\tau+t-1)}{B(\nu,\tau)}$$

i.e. a result similar to the one we would obtain in the binomial case, which remind us that the likelihood of binomial and geometric are the same. As previously stated after t - 1 unsuccessful cycles for a woman, the posterior distribution $f(\lambda|T>t-1)$ is beta$(\nu,\tau+(t-1))$ and thus the marginal hazard μ_t say is

$$E\left(Y_t\middle|Y_{(t-1)}\right) = \mu_t = \frac{\nu}{\nu+\tau+(t-1)}$$

and $Var\left(Y_t\middle|Y_{(t-1)}\right) = \mu_t(1-\mu_t)$

Weinberg and Gladen pointed out the fact that the model implies a linear regression model for the reciprocal of the marginal hazard against cycle t.

$$\mu_t = \frac{\nu}{\nu+\tau+(t-1)}$$

becomes

$$\frac{1}{\mu_t} = \frac{\nu + \tau + (t-1)}{\nu} = \frac{1}{\mu} + \frac{1}{\nu}(t-1)$$

A generalized linear model (GLM) for the marginal response given $(t - 1)$ as explanatory variable may be defined to use a GLIM - type algorithm.

$$g(\mu) = \eta$$

$$\eta = \beta_0 + \beta_1(t-1)$$

$$g(\,) \quad \text{reciprocal link}$$

$$V(\mu) = \mu(1-\mu)$$

Where $g()$ is the link function, η the linear component, $V(\mu)$ the variance function, and β_0 and β_1 respectively an intercept and a slope in the reciprocal scale. The model is fitted to the marginal pregnancy rates.

Note : SAS PROC GENMOD is used to fit the model :

```
proc genmod;
model success/trials=cymin1/dist=bin link=pow(-1);
where attempt=1;
run;
```

"success" and "trials" are respectively observed number of successes per cycle and number of trials, i.e. inseminations, as presented in the Table 6;

"cymin1" represents *t-1* where *t* is the cycle rank;

This analysis concerns only the first series ("attempt") of cycles of the women.

We obtain the results presented in Table 25.

Parameter	Estimate	SE
Intercept	8.35	0.43
Slope	0.71	0.13

Table 25 Estimates of the regression parameter of the reciprocal of the marginal hazards.

Figure 8 presents these results graphically

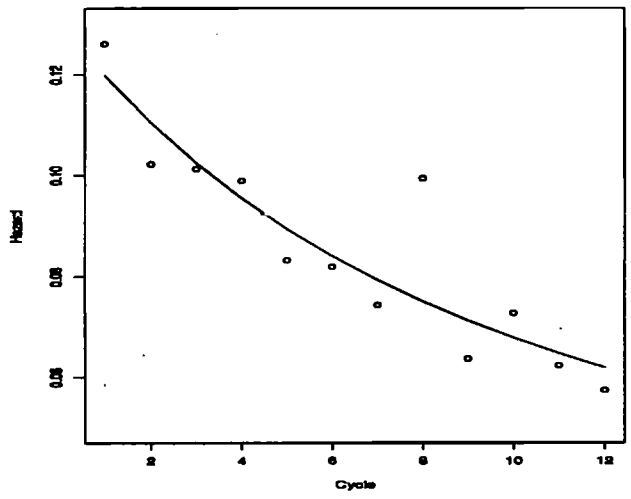


Figure 8 Observed and fitted marginal hazards in first attempts to conceive

From Table 25 $\hat{v} = \frac{1}{\hat{\beta}_1} = 1.4$ and $\hat{\tau} = \frac{\hat{\beta}_0 - 1}{\hat{\beta}_1} = 10.3$ and thus the first two moments of $f(\lambda)$ are respectively 0.12 and 0.008 (standard deviation 0.090). These results may be compared with those obtained above having defined $f(\lambda)$ only through the first two moments respectively, 0.13 and 0.02 (standard deviation 0.145). They are not equal mainly because the latter where obtained using only the results of the two first cycle, with a rapid decrease of hazards. The likelihood estimation of the variance is lower.

As stated in Chapter 3 there are two equivalent ways to write the likelihood of censored data. We have modelled the delays, we will now, for illustration, write the likelihood as the product of conditional contributions of the whole study group over successive cycles :

It is identical to a likelihood arising from a sequence of 12 binomial trials, one at each cycle, where the binomial parameters are the number of women at risk and μ_t.

Likelihood estimate may be obtained using an algorithm composed of three parts [i] the binomial likelihood is specified in terms of μ_t, [ii] μ_t is expressed in terms of ν and τ and finally [iii] a minimization algorithm is used.

Note : S PLUS can be used to fit the model by maximum likelihood. The progam is shown below.

```
note μt is expressed in terms of ν and τ
beta.geom <- function(tr, log.nu, log.tau) {
   nu <- exp(log.nu)
   tau <- exp(log.tau)
   nu/(nu+tau+tr-1)
}

note the binomial likelihood is a specification in terms of  μt
note minimization function provided by S plus  ms()

bg <- ms(~ 2* (D*log(D/N) + (N-D)*log(1-D/N)
           - D*log(beta.geom(tr, log.nu, log.tau))
           - (N-D)*log(1-beta.geom(tr, log.nu, log.tau))),
     start = list(log.nu= lns, log.tau= les))
print(summary(bg))
```

Naturally, we obtained the same result as with the previous method.

The beta-binomial distribution

For the particular case of binary response data an early proposal for hierarchical modeling focused on this beta-binomial model (Sheltam, 1948 ; Chatfield, 1970 ; Griffiths, 1973 ; Crowder, 1995).

Recall that we observed r successes among m inseminations with a sperm donor

$$r \sim Binomial(m, \lambda)$$
$$\lambda \sim Beta(\nu, \tau)$$

and thus the marginal probability for r is :

$$f(r) = \int_0^1 f(\lambda)\binom{m}{r}\lambda^r(1-\lambda)^{m-r}\partial\lambda$$

$$= \binom{m}{r}\frac{B(\nu+r, \tau+m-r)}{B(\nu,\tau)}$$

i.e. the beta-binomial distribution whose mean and variance of $f(r)$ are respectively

$$E(r) = mE(\lambda) = m\frac{\nu}{\nu+\tau}$$

and

$$Var(r) = m(m-1)Var(\lambda) + mE(\lambda)(1-E(\lambda)) = m\frac{\nu\tau(m+\nu+\tau)}{(\nu+\tau)^2(1+\nu+\tau)}$$

Likelihood estimate are $\hat{\nu} = 2.37$ and $\hat{\tau} = 22.23$ and thus $\hat{\mu}$ =0.096 and $\hat{Var}(\lambda)$=0.0034. Recall that Williams method results were respectively : 0.096 and 0.0032 . Thus for the donors the moment estimates — QL and PL — are closer to likelihood estimates than it was the case for the women.

Note : Beta-geometric bivariate model

Here we just give the simple way through which a parametric distribution of the probability of conception enters in the bivariate case, as a complementary note to our previous short presentation of that topic. As above, S and T denote the delay until conception in one

attempt and the subsequent attempts for a same woman. Recall that association may be measured by the odds ratios

$$\theta_{st} = \frac{\Pr(S = s, T = t)\Pr(S > s, T > t)}{\Pr(S = s, T > t)\Pr(S > s, T = t)}$$

These conditional probabilities are presented below in a 2*2 table

Pr(S=s,T=t / S>s-1,T>t-1)	Pr(S=s,T>t / S>s-1,T>t-1)	Pr(S=s/ S>s-1,T>t-1)
Pr(S>s,T=t / S>s-1,T>t-1)	Pr(S>s,T>t / S>s-1,T>t-1)	Pr(S>s / S>s-1,T>t-1)
Pr(T=t/ S>s-1,T>t-1)	Pr(T>t / S>s-1,T>t-1)	$\frac{1}{\Pr(S > s-1, T > t-1)}$

These probabilities are obtained using the following approach :

$$\begin{aligned}
\Pr(S>s,T>t) &= \int_{\lambda} \Pr(S>s,T>t|\lambda) f(\lambda)\partial\lambda \\
&= E\left[(1-\lambda)^{s}(1-\lambda)^{t}\right] \\
&= E\left[(1-\lambda)^{s+t}\right] \\
&= \int (1-\lambda)^{s+t} f(\lambda)\partial\lambda \\
&= \int (1-\lambda)^{s+t} \frac{1}{B(\nu,\tau)} \lambda^{\nu-1}(1-\lambda)^{\tau-1}\partial\lambda \\
&= \frac{1}{B(\nu,\tau)} \int \lambda^{\nu-1}(1-\lambda)^{\tau+s+t-1}\partial\lambda \\
&= \frac{B(\nu,\tau+s+t)}{B(\nu,\tau)}
\end{aligned}$$

Similar formulae being used for other components of θ_{st} expected values for each "cell" s,t may be obtained. These formulae provide a way to calculate predicted values of θ_{st}.

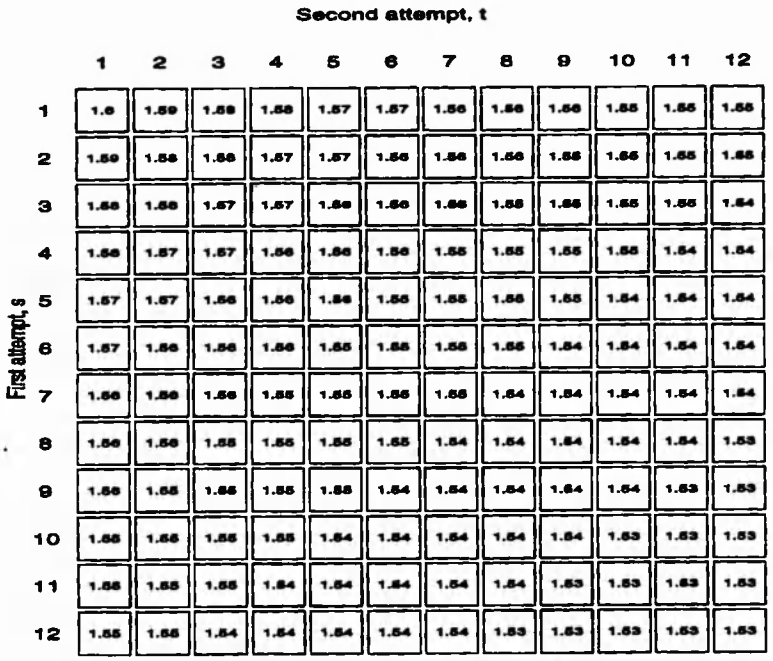


Figure 9 Predicted values of θ_{st} based on the beta-geometric bivariate model. ν and τ being estimated for the 432 women observed at least during two attempts.

Figure 9 shows the values of θ_{st} suggested by the beta-geometric bivariate model. For practical purposes, θ_{st} may by regarded as constant, that is to say that the association between the hazards (probability of success) on two successive attempts of each woman does not depend on the cycles rank.

4. The specification of a mixed model

In Chapter 3 marginal regression models for discrete time censored data were presented : in these models observed covariates were introduced and considered to have *fixed* effects on the risk of conception (successful insemination).

So far, in the present Chapter, we have considered models in which the probability of conception was considered a *random* effect with a distribution, which could correctly take into account the heterogeneity either among women or among donors but did not provide a natural way to take into account the fixed covariates.

We shall do so in the rest of this Chapter. Now we will consider a model containing both *fixed and random* effects. The expected risk of conception will be allowed to change between subgroups of women (or donors) sharing a common set of covariates (fixed effects). Moreover, an added heterogeneity, due to the existence of unobserved covariates (random effects), will be allowed between women (or donors) despite identity of observed covariates.

A question arises naturally at this point : what are the respective "positions" of these two types of effects ? For example, how may complementary log-log model or logistic models presented in Chapter 2, be modified to take account of the overdispersion due to unobserved heterogeneity ?

4.1 Introduction of fixed effects in the beta-geometric model

Recall that the beta-geometric model is based on the following hierarchy : the fecundability of each woman is supposed to be constant and thus the distribution of the delay until conception is geometric. Moreover, the distribution of the fecundability between women is a beta distribution. We have stated that μ_t being the marginal hazard, following the beta-geometric model, we have

$$\frac{1}{\mu_t} = \frac{1}{\mu} + \frac{1}{\nu}(t-1)$$

where μ is E(λ), t is the cycle rank, the distribution of λ among the women being a beta(ν,τ) distribution. Let $\beta_0 = \frac{1}{\mu}$ the intercept on the reciprocal scale and $\beta_1 = \frac{1}{\nu}$ the slope.

Weinberg and Gladen (1986) propose to compare *populations* having separate values of β_0 and β_1 : the reciprocal of the marginal hazards of the populations are thus allowed to vary through the *intercept* and the *slope* component of a linear part of the regression model presented above.

The model of Weinberg and Gladen (1986) is expressed as a parametric model for cycle-specific hazard, μ_t, as follows

$$E(Y_t) = \mu_t$$
$$Var(Y_t) = \mu_t(1-\mu_t)$$

$$\text{and } g(\mu_t) = \frac{1}{\mu_t} = \eta_0 + \eta_1(t-1)$$

where η_0 and η_1 may vary from population to population according to a regression model

$$\eta_0 = \beta_{00} + \beta_{01}x$$

$$\eta_1 = \beta_{10} + \beta_{11}x$$

Table 26 shows the results of this analysis using azoospermia of the husband as covariate.

Model	Parameter	Estimate	SE	Deviance
I	Intercept	8.35	0.43	25.2
	Time	0.71	0.13	
II	Intercept	9.53	0.70	18.9
	Time	0.69	0.13	
	Azoospermia	-1.73	0.71	
III	Intercept	8.88	0.80	17.1
	Time	0.97	0.26	
	Azoospermia	-0.86	0.95	
	Azoospermia * Time	-0.37	0.29	

Table 26 Models for azoospermia.

The introduction of the azoospermia decreases significantly the deviance. There is no further decrease of the deviance when introducing the interaction.

Weinberg's model describes eventually the marginal hazards and the changes in the average conception rate in the population when the covariates are modified. The parameters of the model have a "population averaged" interpretation, they do not describe how the conception probability would be modified for a given woman if she switches from a covariate category to an other. Moreover, they do not have a direct interpretation neither as the log of relative risks nor as log odds ratio.

We can also think of the model as a model for the population distribution of the subject-specific hazard. The parameters ν,τ of this beta distribution are related to Weinberg and Gladen's linear predictors η_0,η_1 by the relationships

$$\nu = \frac{1}{\eta_0} \text{ and } \tau = \frac{\eta_0 - 1}{\eta_1}$$

In terms of the mean and variance of the subject-specific hazard, λ,

$$\text{mean}(\lambda) = \mu_1 = \frac{\nu}{\nu+\tau} = \frac{1}{\eta_0}$$

and

$$\text{Var}(\lambda) = \frac{\nu\tau}{(\nu+\tau)^2(\nu+\tau+1)} = \frac{\eta_1}{\eta_1+\eta_0}$$

This model is rather unattractive. In particular it does not correspond, ever approximatively to a model in which heterogeneity and fixed effects combine additively at the subject level. Such model might be

$$\frac{1}{\lambda} = \eta_0 + \varepsilon$$

so that, to first approximation

$$E(\lambda) \approx \frac{1}{\eta}$$

$$\text{Var}(\lambda) \approx \frac{1}{\eta^4}\text{Var}(\varepsilon)$$

This does not correspond to the Weinberg and Gladen model, unless rather curious assumption are made about the relationship between Var(ε) and covariates.

An alternative approach might be to motivate a model for (ν,τ) using this approximate subject level approach. For example, using a logit model

$$\log\frac{\lambda}{1-\lambda}=\eta+\varepsilon$$

leads to, as a first approximation

$$E(\lambda)=\mu\approx\frac{e^{\eta}}{1+e^{\eta}}$$

and

$$Var(\lambda)\approx\mu^2(1-\mu)^2Var(\varepsilon)$$

Assuming Var(ε)=Constant, σ^2 say, allows the parameters (ν,τ) to be expressed as simple functions of η and σ^2. This approach does not seem to have been explored.

4.2 Introduction of fixed effects in overdispersion models for binary data

In Section 3.1 we have presented a first model proposed by Williams (1982). As an alternative model, Williams proposed to model directly the variance of the probability of success on the logit scale instead of doing it on the probability scale as it was the case for his first approach and for the modelling through the beta distribution.

$$\text{logit}(\lambda_k)=x_k{}^T\beta+b_k$$

$$Var(b)=\phi$$

where $x_k^T\beta$ is the fixed part of the linear predictor, and $b, b_1,...b_k,...b_d$, represents the effect of unobserved characteristics of each of the d donors. The variance of the linear predictor is then independent of its expectation.

The advantage of this approch is that the fixed and the random part of the model add their effects on the same scale, the logit scale. This has been called the "unit treatment additivity" by Cox (1984). It can also be stated that "unit treatment additivity" is obtained if, on the scale where covariates are included additively, the variance due to the unobserved covariates does not depend on these observed covariates. Note that, in introducing covariates on the reciprocal scale in the beta-geometric model, this condition was not fulfilled. The logistic mixed model respecting the unit treatment additivity condition is particularely appropriate if we think that the random covariates are effects of unobserved covariates of the donors (or women).

Up to now we have not specified any distribution for b, that is for logit(λ). As a matter of fact no distribution gives tractable solution for marginals! Later we shall consider that this distribution is normal. At this point it will be easy to introduce both the random effects of the women and donors! This will be discussed in the following Chapters.

At the end of this Chapter we have obtained evidence for heterogeneity among women and donors, which can be considered to some extent as the effect of unobserved characteristics. Modeling the marginal rates provides a first estimation of this heterogeneity.

It appears clearly, however, that in order to estimate correctly the components of the variance due to heterogeneity among women and that due to heterogeneity among donors we need a more complex model incorporating both and meeting the constraint of unit-treatment additivity.

Chapter 5 Unit specific regression models

In the previous chapters we have observed that in this application the correlation structure of our data generated by a double hierarchical design is too complicated to allow regression models for marginal conception rates to be entirely satisfactory.

In this chapter, unit specific regression models for binary data will be presented. With these type of regression models, we are confronted to numerical integrations which renders their estimation difficult. An interesting family of random effect models for discrete survival data will be described. If the random part of the model is limited to only one parameter, this family of model allow the integration to be done analytically. The use of this approach will provide an interesing description of the woman heterogeneity.

1. General presentation of unit specific regression models

Here, each outcome is related to the observed covariates and to a "random effect" shared by all cycles for the same woman. The random effect is included in the linear predictor part of the model. The interpretation of the fixed effect in this model is different from that in the marginal model, since they now represent the effects of the covariates on each woman's probability to conceive. In contrast with marginal models, this approach will also permit the donor random effects to be incorporated in the model in the same way as the recipient effect. Note that, since there is no systematic assignment of donors to recipients, the donor heterogeneity parameter will generally change from cycle to cycle within each woman.

1.1 Generalized linear mixed models

Unit-specific regression models for binary data may be described in the general framework of generalized linear mixed models (GLMMs), i.e. generalized linear models (GLM) which include one or more random effects.

Let y_i $(i = 1, ..., n)$ denote an observed response (result of each menstrual cycle with insemination) assumed to come from a binary process, with λ_i as mean and variance $Var(y_i)$, $\lambda_i(1- \lambda_i)$.

The mean responses are related via a link function, $g(\ )$, the logit or the complementary log-log link, to the elements of a linear predictor η ; the linear predictor is given by a linear regression model involving a fixed part, $x^T\beta$, and a random part, $z^T b$. Formally

$$y_i \sim Bernouilli(\lambda_i)$$

$$g(\lambda_i) = \eta_i$$

$$\eta_i = x_i^T\beta + z_i^T b_i$$

$$\text{where } g(\lambda) = \log\frac{\lambda}{1-\lambda} \text{ for logit link}$$

and

$$= \log\left(-\log\left(1-\lambda\right)\right) \text{ for the complementary log-log}$$

and b are random effects drawn from some probability distribution, $f(b)$ say. It is always possible to consider that the random parameter has fixed location — e.g., zero mean — which will simplify the calculations. The parameters of the distribution of random effects will be denoded by θ.

If a Gaussian distribution is assumed, θ is a *variance*; for example :

$$b \sim N(0, D(\theta)),$$

$$\text{where } D(\theta) = \sigma^2 I \text{ and } \theta = \sigma^2$$

but other distributions may also be used, as will be stated later in this Chapter. For simplicity of notation $D(\theta)$ will be written as D. Z is the design matrix with rows z_i^T ; Z can either be simply a design matrix identifying the clusters (the model is then a "random intercept" model) or contain a direct product of the design matrix and of some or all of the columns of X (the model is in this case a "random intercept and random slope" model (Bryck and Randenbush, 1992 ; Goldstein, 1995). This formulation encompasses situations where the random effects are nested within subjects (multilevel models), and where they are not (more general mixed models).

1.2 Random multiplier : frailty

In the following, depending on the context, the fixed and random effects will be either presented as additive on a linear predictor or multiplicative on an other scale. Recall from Chapter 3 that λ_{it}, the hazard for woman i at cycle t is given, respectively in the logistic model by $\frac{\lambda_{it}}{1-\lambda_{it}} = \frac{\lambda_{0t}}{1-\lambda_{0t}} e^{x_{it}^T \beta}$ and in the complementary log-log model by $-\log(1-\lambda_{it}) = -\log(1-\lambda_{0t}) e^{x_{it}^T \beta}$ i.e. , these models are multiplicative, respectively in $\frac{\lambda}{1-\lambda}$ and in $-\log(1-\lambda)$. The "mixed" version of these same models are now

$$\frac{\lambda_{it}}{1-\lambda_{it}} = \frac{\lambda_{0t}}{1-\lambda_{0t}} \xi_i e^{x_{it}^T \beta}$$

for the logistic model, and

$$-\log(1-\lambda_{it}) = -\log(1-\lambda_{0t})\xi_i e^{x_{it}^T\beta}$$

for the complementary log-log model, where $\xi_i = e^{z_i^T b}$. This unit specific effect ξ, introduced on the multiplicative scale is the commonly called "frailty". Note that on this scale (the frailty scale) the coefficient of variation of the risk is the standard deviation of ξ. Actually, for complementary log-log, for example,

$$\text{var}\left[-\log(1-\lambda)\right] = \left[-\log(1-\lambda_0)e^{x^T\beta}\right]^2 \text{var}(\xi).$$

1.3 Application to our data set

Our data set has a crossed hierarchical structure, each hierarchy with 3 levels (see Figure 1) : menstrual cycles are the level I units ; the cross classification is at level II with respectively :

for the female hierarchy the attempts at level II and the women at level III

and for the male hierarchy the donations at level II and the donors at level III.

Nevertheless it is wise to note that a same donation is not used for all cycles of an attempt and that conversely more than one woman share sperm of each donation.

To fit these data we will use a unit specific regression model. The fixed part of the linear predictor is $X\beta$. All the observed covariates, i.e. all explanatory variables whatever the level (cycle, attempt, woman, donation, donor) may be included in this "fixed" part. The random part of the linear predictor is Zb. Each level of heterogeneity, either associated with the clusters (attempts, women, donations, donor) with the random effect of the covariates, may be taken into account through a random parameter in the random part of the model. The random part related to the clusters may be introduce writing

$$Z = \{A, F, D, M\}$$

$$b = \left\{a^T, f^T, d^T, m^T, \right\}^T$$

where *a, f, d* and *m* are respectively for attempt, female (woman), donation and male (donor) and *A, F, D, M* are "design" matrices assigning cycles to attempts, women, donations and donors, respectively. If, for simplicity of notation, we first ignore attempts and donation, for the *i*th cycle, the model may be written

$$y_i \sim Bernoulli(\lambda_i)$$

$$g(\lambda_i) = \eta_i$$

$$\text{where } \eta = X\beta + Ff + Mm$$

whith *X* representing all covariates related to the menstrual cycle (covariates concerning the woman, this cycle, and the donor), *Ff* the random part corresponding to the woman and *Mm* the random part corresponding to the donor.

2. Inferential process

Three types of inference may be of interest : [i] Inference regarding fixed effects , β, (predictors of fecundability); [ii] inference regarding variance components, θ, (conditional variance of the fecundability, among women or among donors) and [iii] inference regarding the random parameters.

The usual inference strategy is to work in two steps : Estimate β and θ by maximum likelihood after integration $[Data|\beta,\theta] \propto \int [Data|\beta,b][b|\theta]\partial b$ and then estimate the random effects b by empirical Bayes estimates $\hat{b} = E\left(b|Data,\hat{\beta},\hat{\theta}\right)$ i.e. the mean of the distribution

$$\left[b|Data,\beta=\hat{\beta},\theta=\hat{\theta}\right] \propto \left[Data|\beta=\hat{\beta},b\right]\left[b|\theta=\hat{\theta}\right]$$

In some models this likelihood may be obtained in closed form as will be discussed later in this Chapter.

The discussion of Chapter 3 concerning independent censoring must be revisited to take into account of the heterogeneity and the structure of the integrated likelihood. The contribution of each woman to the likelihood depends now on the unobservable parameter b which has to be integrated on

$$L=\int \left\{\prod_t \left[Y_t,C_t|H_{t-1},b\right]\right\}[b]\partial b$$

A factorization under the integration sign shows the respective parts of the event and censoring processes

$$L=\int \left\{\prod_t \left[Y_t|H_{t-1},b\right]\left[C_t|Y_t,H_{t-1},b\right]\right\}[b]\partial b$$

If we assume that the censoring is conditionally independent of the random effect, i.e. $\left[C_t|Y_t,H_{t-1},b\right]=\left[C_t|Y_t,H_{t-1}\right]$:

$$L=\prod_t \left[C_t|Y_t,H_{t-1}\right]\int \left\{\prod_t \left[Y_t|H_{t-1},b\right]\right\}[b]\partial b$$

and, moreover, if the censoring is independent, i.e. $\left[Y_t \middle| H_{t-1}, b\right] = \left[Y_t \middle| Y_{t-1}, X_{t-1}, b\right]$,

then inference based on the maximization of

$$\int \left\{ \prod_t \left[Y_t \middle| Y_{t-1}, X_{t-1}, b\right] \right\} [b] \partial b$$

is consistent. Similar considerations apply when considering the problem of selection for subsequent attempts within the multivariate failure time context. As stated in Chapter 3, interval pregancies violate the independent censoring assumption as does miscarriage, which increases the probability of a subsequent attempt and which is probably related to fecundability — since early conception seems to carry a lower risk of miscarriage than later conception .

However neither of these events are numerous and they will be ignored in subsequent analysis.

3. Discrete time analogues of Hougaard's results

In this section, a first category of mixed models is presented : models offering a closed form of marginals for censored discrete time survival data. Conoway (1990) discussed a *gamma-binomial* model for over-dispersed binomial data providing closed form of marginals for binomial data with covariates. In 1997, Clayton and Ecochard proposed a closely related *gamma-geometric* model for occurrence event time data on a discrete time scale. These models closely parallel the beta-binomial and the beta-geometric models presented in the Chapter 3, but allow time-varying covariates and the introduction of fixed and random effect on a same scale. More generally the complementary log-log mixed models have closed form of marginals for a range of distribution of the frailty parameter.

These results are discrete time analogues of Hougaard's results. In the following we consider only the woman random effect. The introduction of more than one random effect will be postponed until Chapter 7.

Hougaard's results on continuous time scale

As previously stated for survival data on continuous time scale, models for heterogeneity have been proposed, for example by Clayton (1978) and by Vaupel et al (1979), who introduced an unobserved quantity, the frailty, to account for the fact that the prognosis depends usually on unobserved characteristics of the person. The hazard at time t for a person with frailty ξ is assumed to be of the form

$$\lambda(t;\xi) = \xi\lambda(t)$$

The conditional survivor function given frailty is

$$S(t|\xi) = \exp\left(-\xi\int_0^t \lambda(u)\partial u\right) = \exp(-\xi\Lambda(t))$$

where $\Lambda(t)$ is the cumulative hazard. The population — marginal — survivor function $S(t)$ can be written as

$$S(t) = \int \exp(-\xi\Lambda(t))f(\xi)\partial\xi \text{ for all t.}$$

Thus, the survivor function is an analogue of the Laplace transform of $f(\xi)$. Indeed, this Laplace transform is also the *moment generating function* of $f(\xi)$. Several authors, including Vaupel et al (1979), have studied the continuous time model with gamma distributed frailties, which constitute a very convenient family for these models. Hougaard

(1984 ; 1986) has extended these models to the entire class of non-negative exponential families including the frailty parameter, ξ, as a canonical sufficient statistic.

Let $f(\xi)$ denote the density of the frailty distribution. The exponential families we will consider are of the form

$$f(\xi) = \frac{1}{\Phi(\nu,\tau)} \xi^{\nu} m(\xi) e^{-\tau\xi}$$

where m(ξ) is a function of ξ which does not depend on the parameters of the distribution. The gamma distribution is a member of this family having ξ and log(ξ) as sufficient statistics for the parameters ν and τ. The inverse Gaussian distribution has ξ and $1/\xi$ as sufficient statistics and therefore is also a member of this family of distributions. These two distributions will be discussed in more details in this Section. Some closed forms of the MGF are available for them. As a consequence, the survivor function being the MGF of the distribution of ξ, closed forms of the marginal survival function are available.

$$S(t) = \int e^{-\xi\Lambda(t)} \frac{1}{\Phi(\nu,\tau)} \xi^{\nu} m(\xi) e^{-\tau\xi} \partial\xi$$

$$= \frac{1}{\Phi(\nu,\tau)} \int \xi^{\nu} m(\xi) e^{-(\tau+\Lambda(t))\xi} \partial\xi$$

$$= \frac{\Phi(\nu,\tau+\Lambda(t))}{\Phi(\nu,\tau)}$$

Analogues of Hougaard's results on discrete time scale

Clayton and Ecochard (1997) introducing the frailty parameter as a multiplier of $-\log(1-\lambda)$ made a connection between discrete and continuous survival function.

The model relates cycle - and woman - specific hazard for woman i, λ_{it}, to possibly time-dependent covariates x_{it} and random woman effect ξ_i as follows

$$\log\left(-\log(1-\lambda_{it})\right) = x_{it}^T\beta + \log(\xi_i)$$

i.e.

$$-\log(1-\lambda_{it}) = \xi_i e^{x_{it}^T\beta}$$

The ξ_i are assumed i.i.d., drawn from an unknown distribution - the frailty distribution. Dropping the woman subscript, i, and writing $\eta_u = x_u^T\beta$ the linear predictor at cycle u, the conditional probability of surviving until cycle t without conception given frailty ξ is

$$\begin{aligned}\prod_{u=1}^{t}(1-\lambda_u) &= \prod_{u=1}^{t}\exp(-\xi\exp\eta_u) \\ &= \exp\left(-\xi\sum_{u=1}^{t}\exp\eta_u\right)\end{aligned}$$

Note that the unit-treatment additively is maintained : observed covariates and heterogeneity factor are multiplicative on the same scale. Moreover this model can incorporate time-dependent covariates because the frailty component factorizes with the cumulative hazard. The marginal survival probability is given by

$$\int\exp\left(-\xi\sum_{u=1}^{t}\exp\eta_u\right)\partial F(\xi) = M\left(-\sum_{u=1}^{t}\exp\eta_u\right)$$

where $M()$ is the moment generating function (MGF) of the frailty distribution. The likelihood contribution of a woman observed from cycle 1 to cycle t, when she is successful, is given by the marginal distribution function for uncensored observations

$$M\left(-\sum_{u=1}^{t-1}\exp\eta_u\right) - M\left(-\sum_{u=1}^{t}\exp\eta_u\right)$$

while a woman who is unsuccessful for the first t cycles and is censored will contribute the value of the marginal survivor function

$$M\left(-\sum_{u=1}^{t}\exp\eta_u\right)$$

Denoting $\Lambda(t)=-\sum_{u=1}^{t}\exp\eta_u$, the marginal distribution and survivor function are respectively

$$\Pr(T=t)=\frac{\Phi(v,\tau+\Lambda(t-1))}{\Phi(v,\tau)}-\frac{\Phi(v,\tau+\Lambda(t))}{\Phi(v,\tau)}$$

$$\Pr(T>t)=\frac{\Phi(v,\tau+\Lambda(t))}{\Phi(v,\tau)}$$

$\Phi(v,\tau+\Lambda(t))$ being $\int\xi^{v}m(\xi)e^{-\xi(\tau+\Lambda(t))}\partial\xi$

The conditional distribution of ξ is obtained through Bayes rule

$$\Pr(X=\xi|T>t)=\frac{\Pr(T>t|X=\xi)}{\Pr(T>t)}\Pr(X=\xi)$$

and

$$\Pr(X=\xi|T=t)=\frac{\Pr(T>t-1|X=\xi)-\Pr(T>t|X=\xi)}{\Pr(T=t)}\Pr(X=\xi)$$

Thus distribution of ξ among women waiting for conception after t cycles is the same distribution with parameters v and $\tau+\Lambda(t)$.

$$f(\xi|T>t)=\frac{1}{\Phi(v,\tau+\Lambda(t))}\xi^{v}m(\xi)e^{-(\tau+\Lambda(t))\xi}$$

The mean of frailty distribution among conceptive and censored women are respectively

$$E(\xi|T=t)=\frac{\Phi(\nu+1,\tau+\Lambda(t-1))-\Phi(\nu,\tau+\Lambda(t))}{\Phi(\nu,\tau+\Lambda(t-1))-\Phi(\nu,\tau+\Lambda(t))}$$

and

$$E(\xi|T>t)=\frac{\Phi(\nu+1,\tau+\Lambda(t))}{\Phi(\nu,\tau+\Lambda(t))}$$

The gamma and inverse Gaussian distributions

The gamma and inverse Gaussian distributions are members of non-negative exponential families having ξ as canonical parameter

The gamma density is of this form with

$$m(\xi)=\xi^{-1} \qquad \Phi(\nu,\tau)=\Gamma(\nu)\tau^{-\nu}$$

and the inverse Gaussian with

$$\nu=-\frac{1}{2} \qquad m(\xi)=\frac{\psi^{\frac{1}{2}}\exp\left(-\frac{\psi}{\xi}\right)}{\xi\pi^{-\frac{1}{2}}} \qquad \Phi(\nu,\tau)=\exp(-4\psi\tau)^{\frac{1}{2}},\ \tau\geq 0,\psi>0.$$

As the waiting time lengthens there is a progressive selection of the women, the more fertile conceiving earlier. Thus, the marginal hazard decreases. But the selection process also modifies the distribution of heterogeneity. Except for some specific distributions of the frailty — positive stable distributions (Hougaard, 1986) — the population of women becomes progressively more homogeneous : the variance of the probability of success of the woman who have not yet conceived decreases as the waiting time lengthens. This selection process differs according to the distribution of the frailty. For the gamma distribution the mean and variance of the distribution of the risk of conception among women after t failures are, on the frailty scale, respectively $\frac{\nu}{\tau+\Lambda(t)}$ and $\frac{\nu}{[\tau+\Lambda(t)]^2}$

and thus the coefficient of variation is stable, $\nu^{-\frac{1}{2}}$, showing that the relative heterogeneity is independent of the cycle rank. For the inverse Gaussian frailty distribution, the coefficient of variation, $2^{-\frac{1}{2}}\left(\psi(\tau+\Lambda(t))\right)^{-\frac{1}{4}}$ decreases with time thereby making the population more homogeneous.

These differences could be useful to model situations where the gamma geometric model does not seem to fit the data correctly.

Further results for the gamma-geometric model

The gamma distribution for ξ is

$$f(\xi|\nu,\tau) = \frac{\tau^{\nu}}{\Gamma(\nu)}\xi^{\nu-1}e^{-\tau\xi}, 0<\xi<\infty, \nu>0, \tau>0$$

with $E(\xi)=\frac{\nu}{\tau}$ and $Var(\xi)=\frac{\nu}{\tau^2}$ and the corresponding MGF

$$M(s) = \left(1-\frac{1}{\tau}s\right)^{-\nu}$$

The gamma distribution being a member of non-negative exponential families the general formulae presented above may be applied directly and thus the marginal distribution and survivor function are respectively

$$\left(1+\frac{1}{\tau}\Lambda(t-1)\right)^{-\nu}-\left(1+\frac{1}{\tau}\Lambda(t)\right)^{-\nu}$$

and

$$\left(1+\frac{1}{\tau}\Lambda(t)\right)^{-\nu}$$

In the absence of covariates, $-\sum_{u=1}^{t}\exp\eta_u = -t$, the model reduces to a *gamma-geometric discrete survival time distribution.* It follows that

$$\Pr(T=t)=\left(1+\frac{t-1}{\tau}\right)^{-\nu} -\left(1+\frac{t}{\tau}\right)^{-\nu}$$

and

$$\Pr(T>t) =\left(1+\frac{t}{\tau}\right)^{-\nu}$$

Application of the gamma geometric model to AID

Two different approaches may be used to fit this discrete survival model to our data set : either fit the model to marginal hazards, pregnancy rates at each of the successive cycles, using the discrete analogues of Hougaard's models presented above, or consider the cycle of each woman as clustered units, ignoring the cycle rank. We now present the first option. The second will be presented in the next Section.

We will first fit an "intercept" model and then include observed covariates

Simple intercept model, without any covariate

The likelihood contribution for a woman observed from cycle 1 to cycle t is given by the marginal distribution function for uncensored observations $\left(1+\frac{t-1}{\tau}\right)^{-\nu} -\left(1+\frac{t}{\tau}\right)^{-\nu}$ or the marginal survivor function for right-censored observations $\left(1+\frac{t}{\tau}\right)^{-\nu}$. We fit this model to the data of the first attempt, modifying appropriately the algorithm presented for the beta-geometric model and obtain the following results :

$$\hat{\nu}=1.4$$
$$\hat{\tau}=10.5$$

The results concerning heterogeneity may be provided on three different scales for the sake of comparison with results obtained with other methods :

$Var(\xi)$	On the gamma distribution scale
$Var(Log(\xi))$	On the linear component scale
$Var(\lambda)$	On the hazard scale

In order to obtain estimations on these other scales two solutions may be used : either an exact method or an approximate one — the delta method —.

ML estimations of the mean and variance of ξ , i.e. on the frailty scale, are obtained easily, being the mean and variance of $gamma(\nu,\tau)$ and thus calculated respectively as $\frac{\hat{\nu}}{\hat{\tau}}$ and $\frac{\hat{\nu}}{\hat{\tau}^2}$.

On the complementary log-log scale the frailty parameter appears as $\log(\xi)$. The cumulant generating function $K(s)$ is used to generate the first two moments of $\log(\xi)$.

The moment generating function is $M(s)=\int_0^{\infty} e^{s\log\xi}\frac{\tau^{\nu}}{\Gamma(\nu)}\xi^{\nu-1}e^{-\tau\xi}\partial\xi=\frac{\tau^{\nu}}{\tau^{\nu+s}}\frac{\Gamma(\nu+s)}{\Gamma(\nu)}$. Thus $K(s)=-s\log\tau+\log\Gamma(\nu+s)-\log\Gamma(\nu)$ and the mean is $-\log\tau+\Psi(\nu)$ and the variance $\Psi'(\nu)$ where $\Psi(\nu)$ is the derivative of the logarithm of the gamma function and $\Psi'(\nu)$ is the derivative of Ψ . On the hazard scale the variance of λ is obtained as the difference of the 2nd moment of λ, $E(\lambda^2)$,and the square of the first, $E(\lambda)$. Note that $\xi=-\log(1-\lambda)\sim gamma(\nu,\tau)$. Thus $E(\lambda)=\int(1-e^{-\xi})\frac{\tau^{\nu}}{\Gamma(\nu)}\xi^{\nu-1}e^{-\tau\xi}\partial\xi=1-\frac{\tau^{\nu}}{(1+\tau)^{\nu}}$ and $E(\lambda^2)=1-2\frac{\tau^{\nu}}{(1+\tau)^{\nu}}+\frac{\tau^{\nu}}{(2+\tau)^{\nu}}$

The delta-method, sometimes called "propagation of error" method, may also be used, to obtain an approximate variance for example to write the variance on the linear component scale as a function of the variance obtained on the hazard scale. Let λ be the probability of success, drawn from some probability function. Suppose $E(\lambda) = \mu$ and $g(\ )$ denotes, in our case, the link function.

If we want to estimate $g(\lambda)$, a first order Taylor expansion of g about μ would give us $g(\lambda) = g(\mu) + g'(\lambda)(\lambda - \mu) +$ Remainder. For our approximation we forget about the remainder and obtain $E(g(\lambda)) \approx g(\mu)$ and $Var(g(\lambda)) \approx [g'(\lambda)]^2 Var(\lambda)$. For example, if $g(\lambda) = -\log(1-\lambda)$, then $Var(g(\lambda)) \approx \left[\frac{1}{(1-\lambda)}\right]^2 Var(\lambda)$.

Note that the coefficient of variation of λ and *-log(1-λ)* are close, for small λ, as shown Table 27 where the consequent parameter values are displayed. Indeed

$$Coeff.\text{var}(-\log(1-\lambda)) \approx \frac{1}{1-\lambda} Coeff.\text{var}(\lambda)$$

	Frailty scale ξ	Hazard scale λ	Complementary log log scale $\log(\xi)$*
Mean	0.133	0.120	-2.41
Variance	0.013	0.008	1.02
Coefficient of variation	0.857	0.745	-

Table 27 Gamma geometric applied to first attempt AID data.

* The coefficient of variation would not have any signification on this scale, $\log(\xi)$ being not always positive.

Introduction of covariates

As an example and to compare with other models we fit the gamma-geometric model with one covariate, e.g. the presence of an azoospermia of the husband. The maximum likelihood estimates are displayed in Table 28, and the marginal hazards predicted in Figure 10. Note that fixed and random effets are both given on the same, multiplicative — frailty — scale, as multiplier to the basal fecundability — more precisely $-\log(1-\lambda)$ — of women whose husbands are totally sterile.

Parameter	Estimates
No Azoospermia (odds)	0.115
Azoospermia of the husband (odds ratio)	1.25
Donor frailties (Coefficient of variation)	0.831

Table 28 Gamma-geometric model applied to AID data, on first attempt, with azoospermia of the husband as covariate.

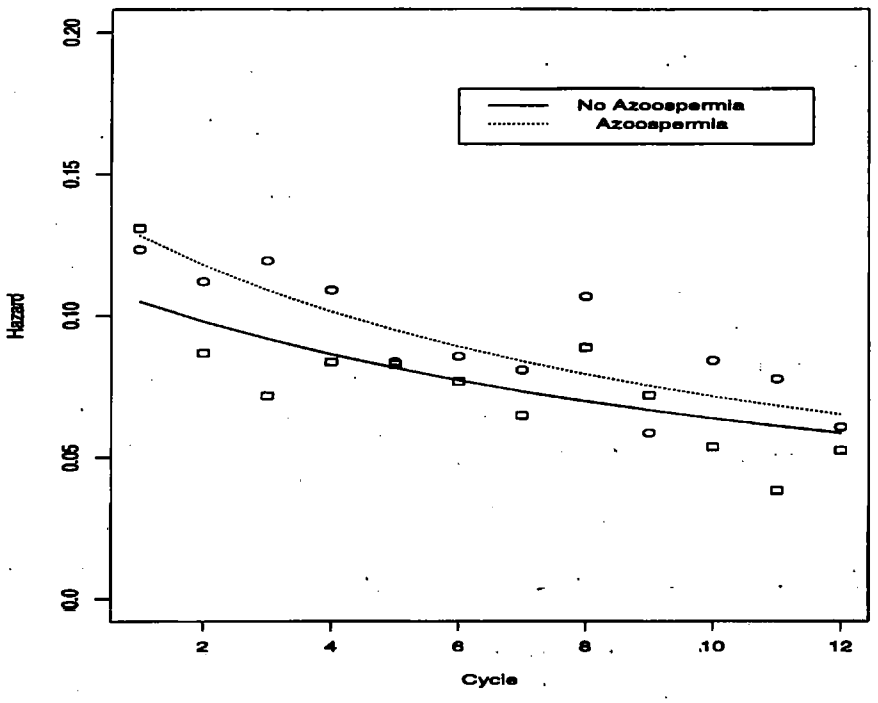


Figure 10 Marginal hazards observed and predicted by gamma geometric model, with (circles) or without (squares) azoospermia of the husband.

We have introduce only one covariate, a time independent one, for illustration. More than one such covariate may be added without any difficulty.

4. Clustered units : use of a Poisson approximation to the binomial likelihood

The data will now be analysed using another option. For this second approach, in a two-level hierarchy, successive cycles of a same woman are considered as unordered units sharing a common frailty. This approach introduces the following Chapters, where the data will be considered as clustered binary data, more than as survival data. Gamma is not a

conjugate of the binomial distribution. Nevertheless a Poisson approximation of the binomial may be used,

leading to a good approximate likelihood method of estimation that may be implemented by use of an algorithm for log-linear Poisson regression model with random effect.

The likelihood contribution for each woman (see Chapter 3) may be written

$$\lambda^{\delta_i}(1-\lambda)^{n_i-\delta_i}$$

where n_i and δ_i are respectively the number of observed cycles and the result of the last one ($\delta_i=1$ for conception, and $\delta_i=0$ for censoring). Let us set

$$\pi_{iu} = -\log(1-\lambda_{iu})$$

where u is the cycle rank and thus, writing without indices for simplicity, the likelihood may be reparameterized using the relations $1-\lambda = e^{-\pi}$

$$\begin{aligned}\lambda &= 1-e^{-\pi} \\ &= \pi - \frac{\pi^2}{2!} + \frac{\pi^3}{3!} - \frac{\pi^4}{4!} \ldots \\ &= \pi\left\{1 - \frac{1}{2}\pi + \frac{1}{6}\pi^2 \ldots\right\} = \pi\left\{1 - \frac{1}{2}\pi + 0(\pi)\right\} \\ &\approx \pi e^{-\frac{\pi}{2}} \quad \text{for small } \lambda, \text{ as is the case for fecundability data}\end{aligned}$$

The likelihood can thus be approximated by $\left(\pi e^{-\frac{\pi}{2}}\right)^d \left(e^{\pi}\right)^{n-d}$ which is equivalent to

$$\pi^d e^{-(n-\frac{d}{2})\pi}$$ i.e., the *Poisson* likelihood

Note that $\pi = \xi e^{\eta}$ and ξ is gamma distributed. This situation leads to an approximate negative binomial likelihood for the parameters of the mixing distribution if covariates are fixed over time (cycles). But time varying covariates may also be included and a log-linear regression be performed. In AID data we have introduced all observed covariates. The algorithm for log-linear Poisson regression model with random effect was written in STATA by D. Clayton and used for our data. The program works by alternating between maximum likelihood estimation of regression coefficients for fixed frailty variance, and estimation of frailty variance for fixed regression coefficients.

Table 29 shows the results.

Parameter	Estimate (s.e.)
Intercept	-2.185(0.062)
Woman :	
Age (woman)	-0.110(0.038)
Azoospermia (husband)	0.094(0.039)
Cycle :	
Insler score	0.242(0.039)
Early insemination	-0.128(0.037)
Late insemination	-0.104(0.033)
Clomiphene citrate	-0.101(0.035)
Donation :	
Sperm count	0.135(0.028)
Sperm mobility	0.184(0.032)
Sperm quality	0.209(0.035)
Heterogeneity :	
Between women	0.666
-2 log likelihood (approximate)	4758.3

Table 29 First attempt. Gamma geometric model. Poisson approximation.

These results will be compared to those obtained with other methods in the following Chapters.

Despite the interest of this model, to date it has proved intractable for more complicated problems involving both hierarchies of AID data. Indeed we could imagine to include one

or more other random effect on the same scale. Yashin and Iachine (1995) have proposed an additive decomposition of frailty components, each gamma distributed such that the total frailty remains gamma distributed. Although an interesting approach, the fact that random and fixed effects are not additive on the same scale is unappealing.

In this Chapter we have introduced unit specific regression models for discrete time survival data. Facing the difficulty to integrate the likelihood function we have limited our analyse to one random effect. Two likelihood methods have been used for inference. A first, based on closed form of marginals obtained in particular way using a gamma geometric model. The introduction of time varying covariates in this first method was not computationally simple. A second method, based on approximation of the likelihood allowed to fit the model to our data including all the covariates. This approach is satisfactory for the female hierarchy with only two levels, the cycles and the women. Nevertheless, AID data set is complicated by the presence of heterogeneity between donors and thus call for a more flexible unit specific regression model. The GLMM with Gaussian random effects will provide an appropriate solution. The next Chapter will present this model and an approximate method for inference (Penalized Quasi-likelihood). PQL will be shown to be a useful and practical way of carrying out preliminary data analysis and Gibbs sampling will provide validation and more accurate results.

Chapter 6 Approximate inference methods for Gaussian random effects models

n the previous Chapters the delay until conception was modelled as geometric. This pproach has allowed the results for one attempt per woman to be described. The last ection of the previous Chapter has considered the successive ovulatory cycles of a same voman as repeated trials sharing a common probability of conception and ignored the ycle rank. Nevertheless the gamma geometric model is limited to one random effect and it s necessary to consider other options which could include both woman and donor random ffects and the corresponding hierarchies. The mixed Gaussian models is a natural way to nclude more than one random effect, and thus it will be described in this Chapter and the ext.

or binary data non linear models have to be used and the estimation procedure becomes ntractable. Approximate inference methods are available for exponential family istributions (Penalized Quasi Likelihood — *PQL* — Breslow and Clayton, 1993). A pecific software for multilevel models, MLn, (Rasbash et al, 1995; Goldstein, 1995) rovides a simple way to fit this model despite the large size of AID data. This Chapter resents the model and PQL. A first application to our data is devoted to an analysis of the ree levels of each hierarchy : woman-attempt-cycles for the female hierarchy and donor-onation-cycle for the male.

1. Inferential methods for Linear Gaussian Mixed models

We first recall the principal aspects of the mixed Gaussian linear model. This Section is intended to provide enough background to understand subsequent application of the generalised linear mixed model for which approximate estimation methods are based on the same principles as those applied to the simple Gaussian linear model.

1.1 Mixed Gaussian linear *model*

The models that underlie the analysis of variance can be viewed as special cases of the general linear model. In this model, y is a response vector; X and Z are matrices of "regressors", β is a vector of unknown parameters, which are called the fixed effects, b is a vector of random effects and e a vector of random errors.

$$y = X\beta + Zb + e.$$

The distribution of y given β is normal with mean $X\beta$ and variance V

$$y \sim N(X\beta, V)$$

V being given by $V = R + ZDZ^T$, where the covariance matrices $D = \mathrm{var}(b)$, $R = \mathrm{var}(e)$, are functions of an unknown parameter vector θ to be estimated from the observations together with fixed effects β. The mixed Gaussian model naturally includes more than one random effect and is an adequate method for describing the effects of unobserved covariates. In contrast a linear combination of independent random effects could not be

Chapter 6 Approximate inference methods for Gaussian random effects models included in a simple way in the models presented in the previous Chapters. The likelihood function viewed as a function of the parameters β and V, with n as sample size, is

$$L(\beta, V|y) = \frac{e^{-\frac{1}{2}(y-X\beta)^T V^{-1}(y-X\beta)}}{(2\pi)^{\frac{1}{2}n}|V|^{\frac{1}{2}}}$$

1.2 Inference

If V is known the ML estimator of β is the generalized least square estimator

$$\hat{\beta} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y$$

The information matrix for $\hat{\beta}$ is $X^T V^{-1} X$ from which we can obtain its precision by matrix inversion. The empirical Bayes estimates of b are obtained setting

$$\hat{b} = DZ^T V^{-1}\left(y - X\hat{\beta}\right)$$

If V is unknown, R and D are estimated from the data as explained below, and introduced in the above equations to get an estimate of β and b.

The substitution of the estimates of fixed effects into the likelihood generates a profile log likelihood function for inference on θ, an unknown parameter definning V.

$$l = -\frac{1}{2}\left(y - X\hat{\beta}\right)^T V^{-1}\left(y - X\hat{\beta}\right) - \frac{1}{2}\log|V|$$

To make degrees-of-freedom adjustment that account for the fact that $\hat{\beta}$ rather than β appears in the quadratic form, but also to protect the estimation of the variance against a mispecification of the fixed part of the model, we use in practice the *REML* (restricted

maximum likelihood) version (Patterson and Thompson, 1971). *REML* is a genuine likelihood for a projection of y into the space orthogonal to that generated by the columns of X. On these basis the *REML* allows the classical likelihood ratio test to be performed. The *REML* equation for Gaussian linear model is

$$l_R = -\frac{1}{2}\left(y - X\hat{\beta}\right)^T V^{-1}\left(y - X\hat{\beta}\right) - \frac{1}{2}\log|V| - \log\left|X^T V^{-1} X\right|$$

Following Harville (1977) we differentiate l_R with respect to the components of θ to obtain estimating equations for the variance components

$$-\frac{1}{2}\left[\left(y - X\hat{\beta}\right)^T V^{-1}\frac{\partial V}{\partial \theta_j}V^{-1}\left(y - X\hat{\beta}\right) - tr\left(P\frac{\partial V}{\partial \theta_j}\right)\right] = 0$$

where $P = V^{-1} - V^{-1}X\left(X^T V^{-1} X\right)^{-1} X^T V^{-1}$. The corresponding information matrix J has components

$$J_{jk} = -\frac{1}{2}tr\left(P\frac{\partial V}{\partial \theta_j}P\frac{\partial V}{\partial \theta_k}\right)$$

1.3 Computational aspects

Closed form solutions being rarely available all these computations require the use of iterative numerical algorithms. A Newton-Raphson algorithm or an *EM* algorithm (Dempster et al, 1977; Lindstrom et al, 1988) may be used to optimize either the full (*ML*) or restricted (*REML*) likelihood function. In the case of AID data both these methods were inefficient because they did not take into account the particular structure of our sparse

covariance matrix. The Iterative Generalized Least Squares estimation procedure - *IGLS*- (Goldstein, 1986) provides an elegant solution for hierachilly nested random effect models. A way to estimate the dispersion components is to choose suitable cross-products of residuals and to equate the observed values of these residuals to their expectations as a function of the parameters characterizing the variance. This is the basic idea of *IGLS*. Goldstein (1986 appendix I) showed the equivalence of *ML* and this iterative generalized least square (IGLS) assuming multivariate normality for θ. We shall present this algorithm using the specific notation proposed in Goldstein (1995).

Goldstein (1986) advocated the following iterative scheme :

(1) Estimation of fixed effect using generalized least square (GLS) : conditional on V the *GLS* of β is given by

$$\hat{\beta} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} y$$

The iterative process starts with $V=I$ which corresponds to the ordinary least square.

(2) Conditional on $\hat{\beta}$ the variance components are estimated using the square of the residuals as response and a matrix derived from the design as regressor.

More formally speaking, conditional on $\hat{\beta}$,

$$vech\left(\tilde{Y}\tilde{Y}^T\right) = Z^{**}\theta$$

where

the residual vector $\tilde{Y} = y - X\hat{\beta}$.

the vector $vech\left(\tilde{Y}\tilde{Y}^T\right)$ is formed by stacking the columns of the lower triangle of the symmetric matrix $\tilde{Y}\tilde{Y}^T$ under one another

Z^{**} is the design matrix relating $vech(\tilde{Y}\tilde{Y}^T)$ to the variance components θ

i.e. $vech\left(\tilde{Y}\tilde{Y}^T\right)$ is treated as a response vector and is regressed, using generalized least squares, upon the columns of a matrix Z^{**}. The corresponding iterative reweighted least square equation is given by :

$$\hat{\theta} = \left(Z^{**T}\left(V^{-1} \otimes V^{-1}\right)Z^{**}\right)^{-1} Z^{**T}\left(V^{-1} \otimes V^{-1}\right)Y^{**}$$

where $Y^{**} = vech\left(\tilde{Y}\tilde{Y}^T\right)$, and $\otimes$ denotes the direct Kronecker product of V^1 by itself, that is an array of matrices which elements $\left[V^{ij}V^{-1}\right]$, V^{ij} being the element of V^1. Goldstein took advantage of the block diagonal structure of the covariance matrx V to avoid its inversion. Each of the block of the expression

$$\left(X^T V^{-1} X\right) \text{ and } \left(X^T V^{-1} y\right),$$

for the fixed effect, and the expression

$$Z^{**T}\left(V^{-1} \otimes V^{-1}\right)Z^{**} \text{ and } Z^{**T}\left(V^{-1} \otimes V^{-1}\right)Y^{**}$$

for the random effect are treated separately needing only the inversion of a low dimension matrix. Then the result are added together to obtain the solution of the relevant equations.

2. An approximate inference method for Generalized Linear Gaussian Mixed models

In the first Section of Chapter 5, we have emphasized the interest of unit specific models to model AID data. The Generalized linear Gaussian mixed model has been presented. But the presentation of the inferential methods was delayed, the second part of the fifth Chapter being devoted to the models having closed form of marginals. In this Chapter and the followings, we use the Generalized linear Gaussian models and study the respective benefits of approximated inference (next Sections and Chapter 7) and of Markov Chain Monte Carlo methods (Chapter 8). In the previous Section we have presented the Gaussian Linear Mixed Model. This linear model does not apply to AID situation where the response is binary. We need a model in which a function of the hazard to conceive is modeled as linearly related to fixed and random effects. The *logit* link function is used.

In order to be able to model more than one random effect we specify the random part as multivariate Gaussian.

Formally,

$$y \sim Bernouilli(\lambda)$$

$$g(\lambda) = \eta$$

$$\eta = X\beta + Zb$$

$$g(\lambda) = \log\frac{\lambda}{1-\lambda}$$

$$b \sim N(0, D(\theta))$$

Two different approximate methods for inference has been proposed (Breslow and Clayton, 1993), Marginal Quasi Likelihood (*MQL*) and Penalized Quasi Likelihood (*PQL*):

The first method aimed at the evaluation of the effects of covariates on marginal pregnancy rate, that is the evaluation of pregnancy rates in specific subgroup. The second method proposes a unit-specific regression model and is based on an approximation of the likelihood. We have emphasized the need for such models for *AID* data, in particular because of the existence of censored observations. Marginal models are appropriate if missing data are missing completely at random (*MCAR*). Stopping rules such as operate in heterogeneous survival data violate this and random effect distributions change with advancing cycle. As has been shown earlier, likelihood methods are applicable in such cases, subject to assumption of independent censoring and censoring not dependent on frailty. Since PQL is an approximation of the likelihood, it is to be prefered.

Thus *PQL* is the method of choice in our case and we will no longer consider *MQL* later in this dissertation.

2.1 Motivation of PQL criterion

To obtain the unconditional likelihood we need to multiply the likelihood by the density of b and to integrate out the random effects :

$$L(\beta,\theta) = \int L(\beta|b)\Phi(b;\theta)\partial b$$

$$= |D|^{-\frac{1}{2}} \int \exp\left(-K(\beta,b,\theta)\right)\partial b$$

where $K(\beta,b,\theta)$ is $-\log L(\beta|b)+\frac{1}{2}b^T D(\theta)b$

A two-step approximation leads to the penalized likelihood for fixed effects

First, writing $\log L(\beta|b)$ as its quadratic expension around the value $\tilde{b}$ of b which minimizes $-\log L(\beta|b)+\frac{1}{2}b^T D(\theta)b$ we can write

$$L(\beta,\theta) \approx |D|^{-\frac{1}{2}} \left|K''(\tilde{b})\right|^{-\frac{1}{2}} \exp\left\{-K(\tilde{b})\right\}$$

therefore, taking account of the fact that $K''(b) \approx Z^T WZ + D^{-1}$

$$\log L(\beta,\theta) \approx -\frac{1}{2}\log\left|I + Z^T WZD(\theta)\right| + \log L\left(\beta|\tilde{b}\right) - \frac{1}{2}\tilde{b}^T D(\theta)^{-1}\tilde{b}$$

where W is the diagonal matrix with diagonal terms being the *GLM* iterated weights, $\lambda(1-\lambda)$ for the logit link case.

Assuming in a second step that the GLM iterative weights vary slowly as a function of the mean, we ignore the first term in the above expression we then have to maximize a penalized likelihood for β given θ (Breslow and Clayton, 1993; Green ,1987)

$$\log L(\beta,b) - \frac{1}{2}b^T D^{-1} b$$

Differentiation with respect to β and b leads, in the logit link case, to the following score equations for the mean parameters :

$$X^T(y-\lambda) = 0$$

$$Z^T(y-\lambda) = D^{-1}b$$

These two sets of score equations will be used respectively to obtain an estimation of the fixed parameters β and the prediction of b.

2.2 Fisher scoring

Green (1987) developed the Fisher scoring algorithm for solutions of these two equations as an Iterated Weighted Least Square problem involving a working dependent variable Y and a weight matrix W that are updated at each iteration. Let us define the working vector Y

to have components $Y_i = \eta_i^b + (y_i - \lambda_i^b) g'(\lambda_i^b)$, where η_i^b is the ith component of the linear predictor including the fixed and random effects, λ_i^b the corresponding hazard and g() the link function.

The solution to our estimating equations via Fisher scoring is equivalent to estimating iteratively β, given by

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

where $V = W^{-1} + ZDZ^T$, and then empirical Bayes estimates of b, obtained setting

$$\hat{b} = DZ^T V^{-1}(Y - X\hat{\beta})$$

It is wise to point out the similarity between these estimating equations and those being used for the linear mixed model presented in the previous Section. The difference consists of the change from y, the response vector, to Y, a working vector, with the corresponding modification of the variance matrix V, in order to incorporate the GLM iterated weights.

2.3 Variance components

The variance components, θ, are estimated through a modified profile likelihood. Indeed some further approximations (Breslow and Clayton 1993) motivate the use of standard estimating equations for variance component written in terms of the working vector Y and the iterated weights W.

The estimating equations according to Harville's can be written

$$-\frac{1}{2}\left[(Y - X\hat{\beta})^T V^{-1} \frac{\partial V}{\partial \theta} V^{-1}(Y - X\hat{\beta}) - tr\left(P \frac{\partial V}{\partial \theta}\right)\right] = 0$$

where $P = V^{-1} - V^{-1}X\left(X'V^{-1}X\right)^{-1}X'V^{-1}$. Again there is a close similarity with the linear case, the modifiction concerning the working vector and the variance.

2.4 Utilisation of IGLS algorithm for PQL

The iterative IGLS algorithm for logistic mixed models uses the same iterative scheme as previously presented for the linear model, but now apply in terms of the *working vector Y* and iterated weights *W*. The estimate of the vector of fixed effects, β, is given by

$$\hat{\beta} = \left(X^T V^{-1} X\right)^{-1} X^T V^{-1} Y$$

and variance components are obtained solving iteratively the following equation

$$\left(Z^{**T}\left(V^{-1} \otimes V^{-1}\right)Z^{**}\right)^{-1} Z^{**T}\left(V^{-1} \otimes V^{-1}\right)Y^{**}$$

where Y is the working vector, Z^{**} the corresponding design vector, Y^{**} the dependent vectors as defined Section 1, and V the current estimates of the covariance matrix.

Note: The "second order approximation" (Goldstein, 1995)

The Fisher scoring algorithms presented above, whose use to estimate the parameters of the logistic mixed model was justified by the fact that they provides estimates maximizing the penalized likelihood (Breslow and Clayton, 1993), may also be obtained through a first order approximation of the model function around the predicted value (Goldstein 1991). Using the fact that this linearization of the model around the predicted value may be

replaced by a more accurate, quadratic, approximation, Goldstein proposed a "second order approximation" method to correct for an eventual bias in the estimations obtained using *PQL*. This interesting discussion will be pursued in the last Chapter as a solution to correct the bias of *PQL*. Nevertheless *PQL* being approximately *ML* is prefered as a first approach. The second order approximation has not, up to now, been shown to be based on any obvious approximation to the likelihood. Although we would speculate that the resilience of *PQL* to data dependent stopping rules will remain, the lack of an objective measure of fit can be a disadvantage in some settings, as we shall see below.

3. A first example of use of PQL for AID data using *MLn*

MLn (Rasbash et al, 1995) is widely used software for multilevel modeling. This Section provides a first reason for using it in the present context. The next Chapter will show other useful features of this software. In this Section we present a first model including two random effects to account for the three-level nature of the female hierarchy.We do not limit the analysis to the first attempt per woman.

Let $(y_1,...,y_n)$ represent the binary responses after each of the n cycles of insemination, where $y_i = 1$ if the ith cycle of insemination ends with a pregnancy, and $y_i = 0$ if not, with $i = 1, ..., n$.

We fit a three level logit model, obtained by assuming that, conditional on the fixed effects, on the woman random effect f, and on the attempt random effect a, the results y_i, of each cycle are independent Bernouilli random variables with probabilities $\lambda_i = \Pr\{y_i = 1\}$ satisfying

$$\text{logit}(\lambda) = X\beta + Ff + Aa$$

where λ is a vector of elements $\{\lambda_i\}$, $X\beta$ is the fixed part of the linear predictor, F is a design matrix for the woman (or : female) random effect f and A is a design matrix for the attempts random effect a. We assume that f have a normal distribution, $N(0,\theta_f I)$ and a a normal distribution, $N(0,\theta_a I)$. The length of f is the number of women (1901) and the length of a is the number of attempts (2437).

Note: The block diagonal structure of V

Following Goldstein (1986), and Goldstein (1991) the block diagonal structure will be easily presented in the case of a this hierarchical model with three levels. Writing Y for the working vector and $\tilde{Y} = Y - X\hat{\beta}$ for the vector containing the random variables, we have

$$E\left(\tilde{Y}\tilde{Y}^T\right) = V$$

which is an $(n \times n)$ matrix where n is the sample size (that is, the number of cycles), and V is constructed in the following recursive fashion. Let V_2 have the following block-diagonal structure:

$$V_2 = \oplus_j \left(V_{1(2)j} + V_{2(2)j}\right) = V_{1(2)} + V_{2(2)}$$

where $\oplus$ denotes the direct sum and $V_{1(2)j} = \theta_e I_{n_j}$ and $V_{2(2)j} = \theta_a J_{n_j}$ with n_j as the number of observed cycles of the jth attempt, J_{n_j} as an $\left(n_j \times n_j\right)$ matrix of ones, θ_e as the variance within attempts and θ_a as the variance between attempts. The V matrices for higher level is constructed in a similar recursive fashion:

$$V = V_2 + Z^{(3)}\theta_f Z^{(3)T}$$

where $Z^{(3)}$ is in our example the design matrix for the women random coefficients and θ_f the variance between women. For each woman a block takes place on the diagonal of V. Figure 11 and Figure 12 provide its schematic aspect.

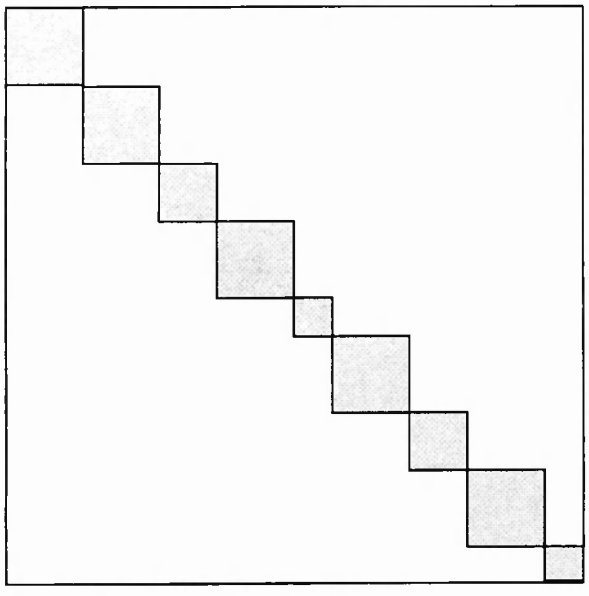


Figure 11 Schematic aspect of V_3. Each square of various size on the diagonal represent a woma

i.e. a block at level III. Outside these blocks the matrix is filled with 0s.

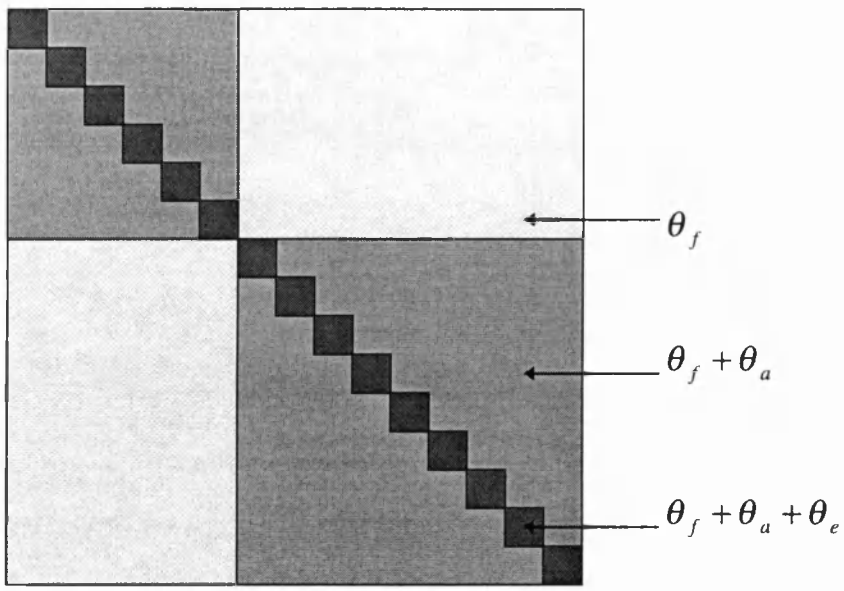
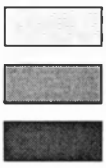


Figure 12 Representation of one block. Example of a woman, with two attempts, with respectively 6 and 9 cycles



We now present precisely how we specify the model using MLn.

Preparation of the data

The data are read by *ML*n from an ASCII file. Each line corresponds to a cycle, identified by identifiers for units at each level : level I, cycles; level II, attempts; level III women. The data are first *sorted* with the level III and level II as major sort keys.

Response variable and hierarchy

```
resp "PREGNANT"
iden 1 "CYCLEID"
iden 2 "WOMEN"
iden 3 "ATTEMPTS"
expl  "ONE1" "ONE2" "ONE3"
```

where "ONE1" is for a column of one to initiate the heterogeneity at level I, "ONE2" is for a column of one to indicate a random intercept at level II and "ONE3" is for a column of one to indicate a random intercept at level III.

```
setv 1  "ONE1"
setv 2  "ONE2"
setv 3  "ONE3"
```

to inform that there are three variance components, one for level I (between cycles) and the others for level II (between attempts) and level III (between women). MLn informs the user of the model setting :

```
IDENtifying codes : 1-CYCLEID, 2-ATTEMPS 3-WOMEN
LEVEL 3 RPM
         ONE3
CONS     1
LEVEL 2 RPM
         ONE2
CONS     1
LEVEL 1 RPM
         ONE1
BCONS    1
```

Until now the model is a Gaussian linear mixed model

$$\lambda = b_{30} + b_{20} + b_{10}$$

where λ represents the hazard and b_{30}, b_{20} and b_{10} denote respectively the random intercept at woman's level, attempt's level and the error at the cycle level.

Logistic link

The modifications of the procedure needed for non-linear models are obtained using the existing MACRO language of *MLn* : these macros transform adequately between each iteration the response variable and the explanatory variables.

Note that Goldstein uses a slightly different parameterization than that presented above. This justifies the need to use "working" explanatory variables (original covariates multiplied by the first derivative of the link function.

PQL and *REML* options proposed by the package the following procedures are choosed

```
set B12 1              note  PQL
meth 0                 note  REML (or : RIGLS)
```

MLn informs the user of the model setting :

```
                        NON-LINEAR SETTINGS
                        ===================
LINK FUNCTION(B10)        : LOGIT(0)
APPROXIMATION(B11)        : FIRST ORDER(1)
NINLINEAR PREDICTION(B12): FIXED+RESIDUALS:PQL(1)
VARIANCE FUNCTION(B14)    : DISTRIBUTIONAL(0)
DATA STRUCTURE(B15)       : UNIVARIATE ANALYSIS(1)
MIXED RESPONSE(B16)       : NO(0)
```

The underlying *logistic* mixed model equation is

$$\text{logit}(\lambda) = b_{30} + b_{20} + b_{10}$$

Introduction of covariates in the model

Whatever be the level of the fixed covariates they have a *comparable place in the linear predictor.*

```
expl 1 "azoocent" "inscent"
```

where `"azoocent"`, azoospermia of the husband is a characteristic of the woman (level III) and `"inscent"` the insler score (level I), characteristic of the cycle. This equivalence of the place of covariates whatever be their level is easily understood. We recall that the underlying mixed model equation is

$$\text{logit}(\lambda) = x_1\beta_1 + x_2\beta_2 + b_{30} + b_{20} + b_{10}$$

where λ represents the hazard, x_1 and x_2 correspond to observed covariates at level I (cycle) and level III (woman), and β_1 and β_2 are the corresponding fixed coefficients, and b_{30}, b_{20} and b_{10} denote respectively the random intercept at the woman's level, attempt's level and the heterogeneity at the cycle level.

Practical aspects

MLn was programmed for economical use of *RAM* (Goldstein, 1986, Appendix 2). *ML*n evaluates elements block by block, as presented in Section 1, (a block size is fixed by the number of unit per cluster) and sums the results, saving time and memory spaces.

Standard variance component procedures (e.g. SAS Proc Mixed) incorporate *ML* estimation for normal response models. But this procedure does not exploit the diagonal structure of the variance matrix in hierarchical model. It does not use neither a specific algorithm for sparse matrices. Thus we did not succeed in implementing the mixed model for AID data in SAS (Macro Glimmix, 1992) : Despite the reduction of the dataset to two third of the cycles a Vax machine 3 600 was not able to converge within one week.

MLn was really successful with our data sets : a few minutes were enough to fit this three level model with 1901 women, 2437 attempts and a total of 12100 level 1 units (cycles), on a 75 MHz Pentium PC.

Results

Mln provides the following results :

PARAMETER	ESTIMATE	S. ERROR(U)	PREV. ESTIMATE
INTERCEP	-2.163	0.03658	-2.163
AZOOCENT	0.0815	0.03637	0.0815
INSCENT	0.2805	0.03918	0.2805

rand

LEV.	PARAMETER	(NCONV)	ESTIMATE	S. ERROR(U)	PREV. ESTIM	CORR.
3	ONE3 /ONE3	(1)	0.5207	0.07145	0.519	1
2	ONE2 /ONE2	(8)	0	0	0	1
1	ONE1 /ONE1	(10)	1	0	1	1

In may be pointed out that the heterogeneity between attempts is very low, presented as "0". This aspect will be discuss in Chapter 7. Since the variance at the base level, the cycle

level, was supposed to be the binomial variance the scale factor is set to 1 in the programme option as can be read on the results.

4. First analyses using the *PQL* approach

This Section presents the results of a separate analysis of female and male hierarchy including all three levels - woman-attempt-cycles for the female hierarchy and donor-donation-cycle for the male. We shall also examine how the introduction of the fixed effect in the model modifies the variance components.

Female hierarchy

Let us first fit a simple intercept model for first, second, and subsequent attempts, successively; the results of the estimation of the variance of the probability of conception among women are given in Table 30.

Attempt	variance on the logit scale, θ_f
First	0.43
Second	0.37
Subsequent	0.40

Table 30 Women at level II. No covariates

These first estimations of the variance on the logit scale obtained using *PQL* (Table 30) are lower than ML estimates obtained using the gamma geometric model presented in the previous Chapter, as shown Table 31.

Attempt	variance on the complementary log-log scale
First	0.74
Second	0.49
Subsequent	0.89

Table 31 Gamma geometric model. ML estimates.

The difference between these two sets of results are not entirely explained by the difference of models : the logitic and complementary log-log links are quite similar for small probability of conception per cycle. The bell-shaped gamma distribution is not so different from the Normal distribution to explain such a discrepancy. Therefore considering that the ML estimates of the gamma geometric model have good properties we must also accept that the PQL underestimate the variance (Breslow and Lin, 1995).

Table 32 presents the results of an analysis including all the observed covariates

Parameter	Estimate (s.e.)
Intercept	-2.237(0.039)
Woman :	
Age (woman)	-0.105(0.037)
Azoospermia (husband)	0.080(0.037)
Cycle :	
Insler score	0.260(0.039)
Early insemination	-0.137(0.038)
Late insemination	-0.084(0.033)
Clomiphene citrate	-0.104(0.036)
Donation :	
Sperm count	0.140(0.030)
Sperm motility	0.175(0.033)
Sperm quality	0.249(0.036)
Heterogeneity :	
Between women	0.569(0.074)
Between attempts	0

Table 32 Complete data. Female hierarchy

These results obtained from the whole dataset and from a model including the three level of female hierarchy does not show any heterogeneity among attempts. This point will be discussed in the next Chapter. The variance of the probability of conception among women, estimated on the logit scale is no longer substantially different from the one obtained using a maximum likelihood method of estimation and the gamma geometric model (0.666, see Table 29; estimation based only on the cycles of the first attempt). The

estimates of the fixed effects are also close to those obtained using a crude marginal model (Chapter 3, Table 20) and those obtained using the gamma geometric (Chapter 5, Table 29) for the data of the first attempt. Using the same data set, but without including the covariates in the model the estimate of the variance between women decreases to 0.520 (0.071). Most of the fixed effect we observe are at the level I (ovulatory cycle) and their distribution among women vary

extremely : this observation explains at least partly the absence of reduction of the variance between women after the introduction of fixed effects.

Donor hierarchy

The estimated variance of the probability of conception among donor (0.387) is very close to the one obtained in previous Chapters (beta-binomial model, ML estimate, estimated variance on the logit scale : 0.452). The PQL bias is therefore small.

Table 33 presents the results of an analysis including all the observed covariates. In this model we introduce the cycle rank and a dummy variable for attempts (0 if first, 1 if subsequent). This introduction of cycle rank and attempt rank give us a way to model marginally (population-averaged) with respect to women, but conditionally (subject-specific) with respect to the donors.

Parameter	Estimate (s.e.)
Intercept	-2.067(0.077)
Woman :	
Age (woman)	-0.116(0.033)
Azoospermia (husband)	0.059(0.032)
Attempt:	
Subsequent	0.266(0.077)
Cycle rank:	
2	-0.312(0.103)
3	-0.276(0.107)
4	-0.320(0.115)
5	-0.474(0.126)
6	-0.500(0.135)
7	-0.548(0.149)
8	-0.307(0.144)
9	-0.676(0.175)
10	-0.676(0.189)
11	-0.918(0.217)
12	-0.960(0.238)
Cycle :	
Insler score	0.253(0.038)
Early insemination	-0.139(0.038)
Late insemination	-0.078(0.033)
Clomiphene citrate	-0.092(0.034)
Donation :	
Sperm count	0.131(0.039)
Sperm motility	0.173(0.042)
Sperm quality	0.201(0.042)
Heterogeneity :	
Between donors	0.230(0.046)
Between donations	0.031(0.047)

Table 33 Complete data.Male hierarchy

At the intermediate level of the hierarchy the variance between donations is estimated as close to zero. This suggest that donor more than donations are responsible for the success of a trial, or that all the covariates explain the variance of the donation. This will be discussed in the next Chapter. There is in addition a striking difference between these results and those obtained for the woman, which is the sharp decrease of the estimated variance when introducing the fixed effects (from 0.387 to 0.230).

Next Chapter will be devoted to the discussion of this complex variance structure and on its effect on the estimation of the fixed and random effect. Nevertheless these analysise shall be possible only after the introduction of the crossed hierarchical structure.

Chapter 7 Extended *MLn* analysis

If we want to take into account simultaneously the male and female factors in a statistical model we are confronted with a crossed hierarchical structure since both factors are exerting their effect on the probability of conceiving at a given cycle through several levels of their respective hierarchy (cycle, attempt, woman; cycle, donations, donors).
In the previous Chapters we have avoided to take into account simultaneously these two hierarchies. In this Chapter both will be included in the model. A crossed random multilevel logistic model, taking account of this structure shall be presented. Moreover having taken into account properly the correlation between ovulatory cycles, an extended analysis of the dataset will be performed.

1. Crossed random multilevel logistic model

As previously stated the statistical complications arise from the doubly hierarchical structure of the dataset (see Chapter 2, Figure 1 C). This double hierarchy induces a complicated correlation structure and statistical methods which would ignore it will tend to underestimate the variance of the coefficient estimators. There is also considerable interest in the extent of extra variability attributable at all levels and to study the fixed effects taking account of their place in the hierarchies.
To face the problem of crossed design, solutions have been proposed by Raudenbush (1993), Rasbash and Goldstein (1994) and Goldstein (1995). Breslow and Clayton (1993) analysed the salamander mating data reported by Mc Cullagh and Nelder (1989) using a

crossed random multilevel logit model. Insemination with donor sperm dataset is globally of the same type, with 3 levels instead of two: the cycles are the level I units; the cross classification is at level II of the hierarchy assigning in a given "cell" the collection of cycles belonging to the same attempt of the same woman and inseminated with the same donation (Chapter 2, Table 3). Attempts to conceive are themselves nested within women and sperm donations within sperm donors. However, in contrast with the salamander mating data, the insemination data are totally unbalanced. Moreover, the majority of the cells are empty, each woman receiving sperm from at most 12 donors (12 cycles is a maximum per woman); there is rather rarely (about 6 % of the cells) more than two cycles in each "cell", since there is no systematic assignment of donors to recipients.

The computations for models with crossed random effects are laborious. Breslow and Clayton, discussing the analysis of the "salamander data" which involves two sets of 60 crossed random effects concluded that the size of this problem corresponds to the limit of faisability for PQL computation; with larger probelms Markov Chain Monte Carlo (*MCMC*) methods become attractive. In the present case we have 1901 women and 279 donors and the potential random effects of 2437 attempts and 1328 donations. It was therefore nevessary to design a method of computation which can deal properly with such a large dataset.

In the next Chapter an *MCMC* solution will be investigated. In this Chapter we apply the *PQL* approach to the insemination dataset, using an alternating *EM* algorithm (Dempster et al 1977) to approach the serious computational difficulties involved in fitting models with large numbers of crossed random effects (women and donors).

Our solution to these difficulties derives from the fact that simple hierarchical random effect models, as stated in Chapter 6, may be fitted extremely efficiently by exploiting the block diagonal structure of the matrices involved. Thus, random effect models for woman

and attempt effects can be rapidly fitted as can similar models for donor and donation effects. We simply use an EM algorithm which alternates between these two sets of rapid computations.

Model

Let $(Y_1,...,Y_n)$ represent the binary responses to a total of n cycles of insemination, where $Y_i=1$ if the ith cycle of insemination ends with a pregnancy, and $Y_i=0$ if not.

For simplicity of notation we will first ignore attempts and donations and allow for only recipient and donor effects. We propose a crossed random multilevel logit model which assumes that, conditional on a woman random effect f and on a donor random effect m, the result of each insemination , y_i, are independent Bernoulli random variables with probabilities $\lambda_i = pr\{Y_i = 1\}$ satisfying

$$\text{logit}(\lambda) = X\beta + Ff + Mm$$

where λ is a vector of elements $\{\ \lambda_i\ \}$, X is the matrix for the fixed effects β, with rows x_i^t, F is the design matrix for the woman (or : female) random effect f and M is the design matrix for the donor (or : male) random effect m. X represents cycle covariates but also covariates at higher levels, concerning women or donors and donations. We assume that f and m both have a normal distribution, respectively $N(0,\theta_f I)$ and $N(0,\theta_m I)$. The length of f is the number of women, and the length of m is the number of donors.

The conditional mean can be written : $E(y|b) = h(X\beta + Zb)$ where h(.) is the logistic function

and $Z = \{F, M\}$, with rows z_i^t, and $b = \left(f^T, m^T\right)^T$

The integrated likelihood function is used to estimate β, θ_f and θ_m

$$L(\beta,\theta_f,\theta_m)=\prod_i \iint L(y_i|\lambda(f,m)_i)\Phi(\theta_f,\theta_m)dfdm$$

where $L(y_i|\lambda(f,m)_i)$ denotes the conditional likelihood of y_i given its mean and $\Phi(\ )$ the multivariate normal density. As previously stated, its logarithm can be approximated following Breslow and Clayton (1993)

$$l(\beta,\theta_f,\theta_m)\approx -\frac{1}{2}\log\left|I+Z^tWZD\right|+l\left(\beta|\tilde{\lambda}\right)-\frac{1}{2}\tilde{b}\,D^{-1}\,\tilde{b}$$

where $\tilde{b}$ is chosen to maximize the sum of the last two terms, that is to say as the empirical Bayes estimates (*EB*), and $\tilde{\lambda}=\lambda(\tilde{f},\tilde{m})$; $D=D(\theta_f,\theta_m)$ is the diagonal covariance matrix of b and W is the diagonal matrix with diagonal terms $\tilde{\lambda}_i(1-\tilde{\lambda}_i)$, the *GLM* iterated weights.

An algorithm to maximize a "penalized" log likelihood of the form of above expression is the following.

1) Calculate the working vector and weights as in the familiar *GLIM* algorithm, using current *ML* estimates of β and current *EB* estimates of b, and calculating "working" Y-values using a local linearization around λ

$$Y_i=\left(x_i^t\beta+z_i^tb\right)+\frac{1}{\lambda_i(1-\lambda_i)}(y_i-\lambda_i)$$

where λ_i is the fitted value using both fixed and random terms in the model, $\text{logit}(\lambda_i)=x_i^t\beta+z_i^tb$, and

2) Iterate between

2a) Solve the mixed model equations for Gaussian linear mixed models, to yield approximate *ML* estimates of the fixed effects, β, and *EB* estimates of the random effects, b,

2b) Calculate maximum likelihood estimates of the variance components (the parameters which determine the matrix D), again under the assumption that the "working" Y-values are Gaussian. In the mixed logistic model this yields maximum pseudo-likelihood estimates.

For single and nested random effects this algorithm has been implemented efficiently in the *MLn* computer program (Rasbash and Woodhouse, 1995) as presented in Chapter 6.

For crossed random effects, a more general program such as SAS PROC MIXED must be used to implement the algorithm. Unfortunately, however, the matrix inversions required in step (2) were not feasible for a dataset of the size considered here.

An alternative computational method has been proposed by Rasbash and Goldstein (1994), but this is only feasible for small datasets and problems in which the crossing of random effects is rather incomplete so that there are not too many groups of subjects sharing distinct pairs of random effects. This method too was not feasible in the present application. Instead we alternated between two *EM* algorithms.

An alternating EM algorithm

If a feasible method for simultaneous computation of estimates of β and b is not available, the *EM* algorithm can be used. Here the solution is obtained by alternating between solving the equations for β and b assumed known, and solving the equations for b with known β.

In our case, there are two sets of random effects, $b = \left(f^T, m^T\right)^T$. Two *EM* algorithms are possible :

1. Solve for β and m in the M-step, using empirical Bayes estimates of f (E-step) as "offsets", or

2. solve for β and f in the M-step, using empirical Bayes estimates of m as "offsets"(E-step).

In fact we alternate between these algorithms since, by so doing, each "M-step" delivers the empirical Bayes estimates required for the alternative E-step. It is fairly straightforward to show that this algorithm will increase the penalized likelihood at each step. Thus, our alternating *EM* algorithm will, for fixed estimates of the variance components, converge on correct estimates of β, m, and f.

This procedure solves the mixed model equations for the fixed and crossed random effects. The estimating equations for the fixed effects, β, are

$$E\left\{\sum_i W_i\left(Y_i - x_i^t\beta - z_i^t b\right)x_i\right\} = 0$$

where the expectation is taken with respect to the posterior distribution of b. The weights W_i are inversely proportional to the (residual) variance of Y_i the "working" Y-value. Taking the expectation on the left hand side is equivalent to replacing the random effects, b, with empirical Bayes estimates.

Estimation of variance components, θ_f and θ_m, is more problematic. We propose to use the estimates delivered by each half of algorithm. Nevertheless this will only yield the same estimates of variance components as would be obtained from fitting the full model under rather special circumstances. It may easily be shown that for the alternating *EM* algorithm to yield the correct solution, the two sets of random effects must be *a posteriori*

independent of one another. The equation for the *ML* estimate of θ_f, the between-woman variance component, is

$$n_f\theta_f = E\left\{\sum_j (f_j)^2\right\} = \sum_j \left\{E(f_j^2)\right\}$$
$$= \sum_j \left\{\left(E(f_j)\right)^2 + Var(f_j)\right\},$$

where j indexes the women, n_f is the number of women and expectations and variances are taken with respect to the posterior distributions of f. A similar equation holds for the *ML* estimate of θ_m. Our algorithm does not in general deliver estimates satisfying these equations because the term $Var(f_j)$ in the above equation is not estimated as it should. For example, when computing the estimate of θ_f we assume the male effects to be known (at $\tilde{m}$). Thus the posterior variance contribution in the equation for θ_f will be taken as the *conditional* posterior variance $Var(f_i|Data,\beta,m=\tilde{m})$ rather than $Var(f_i|Data,\beta)$. If m,f are *a posteriori* independent, then these two variances will be the same, but this will only be true for balanced designs in which the likelihood for m and f factorizes. In other circumstances our algorithm will not deliver correct estimates.

Figure 2 (Chapter 2) demonstrates that, in this application, the assignment of donor to recipient is not balanced. However, donors and recipients are only associated as a result of the date of insemination; when the data are stratified by calendar year of start of treatment (Figure 3), the association between donor and recipient is largely removed.

We have verified that for AID data the estimates of variance components are nearly unchanged after such stratification, and conclude that use of the algorithm is justified in our case.

Computational aspects

The *MLn* algorithm uses the fact that V has a very special structure for nested random effect models to speed up the computations. While this matrix does not have this structure when both random effect included in the model it would have it if one random effect is fixed to the previous estimate.

For practical reasons, the algorithm proposed by Goldstein is applied to the working vector $\tilde{\lambda}(1-\tilde{\lambda})Y$, the design matrices [X, Z] being replaced by $[\tilde{\lambda}(1-\tilde{\lambda})X, \tilde{\lambda}(1-\tilde{\lambda})Z]$.

The logistic mixed model having being declared as in Chapter 6, the *EM* algorithm is implemented in the following way:

(0) Setting $\hat{m}$ to *0* as initial estimate, run *MLn* estimation for β, θ_f and calculate empirical Bayes estimates for f

```
note Setting m̂ to 0 as initial estimate
note C34 is a column of 0
name  C34 "OFFS"
note run MLn estimation for β,θf
start
note calculate empirical Bayes estimates for f
rcov 0
rlev 2
resi
```

(1) Calculate the working dependent vector Y, and the « working » design matrices

$[\tilde{\lambda}\ (1-\tilde{\lambda}\ )X, \tilde{\lambda}\ (1-\tilde{\lambda}\ )Z]$,

Introduce the current estimation of f as offset

Sort the dataset in donor blocks

Run an *MLn* estimation for β, θ_m and calculate empirical Bayes estimates for m.

```
note The working matrices are updated by MLn macros
note "OFFS" is updated to the value of f.
note The data are sorted in donor blocks
note Run an MLn estimation for β,θm
iden 2 "donor"
start
note calculate empirical Bayes estimates for m : see (0)
```

(2) Calculate the working dependent vector Y and of the « working » design matrices

$\tilde{\lambda}\ (1-\tilde{\lambda}\ )X$ and $\tilde{\lambda}\ (1-\tilde{\lambda}\ )Z$,

Introduce the current estimation of m as offset

Sort the dataset in women blocks

Run an *MLn* estimation for β, θ_f and calculate empirical Bayes estimates for f.

Iterate between (1) and (2) until the convergence of the estimates of $\beta, \theta_f, \theta_m$

Note that the two *MLn* fits yield different standard errors at convergence. Both of these are incorrect since each ignores the fact that one set of random effects are estimated. The standard errors shown are based on the final *MLn* analysis which fixes the donor random effects and, as stated above, are misleading. In the case of linear least squares, exact

distributional results for the ordinary least square estimator may be derived and this theory may be used to construct exact confidence intervals as long as the assumption of normality holds. Even if the response is not normally distributed the estimator remains unbiased. In the linear model, $\hat{\beta}$ and $\hat{\theta}$ are asymptotically orthogonal so that $\text{var}\left(\hat{\beta}|\hat{\theta}\right)$ converges to $\text{var}\left(\hat{\beta}|\theta\right)$. In non linear models, $\hat{\beta}$ and $\hat{\theta}$ are asymptotically correlated, and hence inferences about $\hat{\beta}$ must take into account uncertainty in $\hat{\theta}$. In the non linear model, it is no longer possible to obtain exact results even under the normal distribution with constant variance. When calculating estimation of fixed effects and empirical Bayes estimates of random effects, we make the assumption that the variance components are known. This strategy is acceptable for point estimation, but not for estimation of the variance of points estimators.

Significance of covariate effects is better assessed by tests based on (approximate) likelihood ratios. Excluding successively each parameter in turn, we reestimate the others and calculate the maximized log-likelihood, having constrained the variance components to their estimated values derived from the most complex model (McCullagh and Nelder 1989 p 91). The likelihood ratio test statistics are formed by calculating changes in (-2 log-likelihood).

2. Results

We first fit a full intercept model, with three levels for female hierarchy (cycle, attempts, women) and three for the male hierarchy (cycle, donations, donors). Table 34 exhibits the results.

Parameter	Estimate (s.e.)
Intercept	-2.274
Heterogeneity :	
Between women	0.446(0.069)
Between attempts	0
Between donors	0.377(0.058)
Between donations	0
-2 log likelihood (approximated)	7 028.0

Table 34 Complete data. Alternating EM algorithm.

Stability of this EM algorithm is obtained very *rapidly* for all the parameters shown in this Table. Four iterations are sufficient. In spite of multiple checks and verifications we do not obtain any evidence for an identifiable heterogeneity between attempts and between donations when both women and donor effects are in the model. From now on we consider that these intermediate levels have no practical implication and are discarded from the model, except in so far as they are relevant to the construction of "compositional" covariates..

Table 35 exhibits the full model with covariates.

Parameter	Estimate (s.e.)
Intercept	-2.303
Woman :	
Age (woman)	-0.106 (0.036)
Azoospermia (husband)	0.080 (0.037)
Cycle :	
Insler score	0.264 (0.039)
Early insemination	-0.137 (0.038)
Late insemination	-0.082 (0.034)
Cloniphene citrate	-0.103 (0.036)
Donation :	
Sperm count	0.130 (0.030)
Sperm motility	0.179 (0.034)
Sperm quality	0.217 (0.036)
Heterogeneity :	
Between women	0.500 (0.072)
Between donors	0.222 (0.043)
-2 log likelihood (approximate)	6 880.6

Table 35 Complete data. Alternating EM algorithm. With covariates.

This table presents the values of the estimated fixed effects and the estimated variance of the random effects concerning the heterogeneity between women and between donors using the logistic mixed model and the double EM algorithm presented above. The results of this analysis do not change appreciably when stratified by year of treatment in an attempt to minimize association between donor and recipient, $\hat{\theta}_f = 0.505$ and $\hat{\theta}_m = 0.226$.

All the tests are based upon comparisons (ratios) of the approximate likelihood. Excluding successively each parameter we estimate again the others and the log-likelihood after having constrained the variances (in both hierarchies) to their estimation derived from the most complex model.

Note that, to obtain the likelihood ratio test in this mixed model we fit the most complex model and then exclude the tested parameter, having constrained the variance to its estimation derived from the most complex model (McCullagh and Nelder, 1989, p 91). Table 36 shows the results of this procedure, including the ratio test statistics formed by substracting the values of -2 log likelihood. We do it in a backward fashion excluding the parameters of smaller size (these are standardized regression coefficients) among the parameters present in the sub-model.

Parameter being dropped out from the model	-2 log likelihood (approximate)	L ratio test
(Full model)	6 880.6	-
Azoospermia (husband)	6 885.2	4.6
Late insemination	6 891.2	6
Clomiphencitrate	6 899.2	8
Age (woman)	6 908.4	9.2
Sperm count	6 919.4	11
Early insemination	6 930.4	11
Sperm motility	6 948.0	17.6
Sperm quality	7 003.2	55.2
Insler score	7 063.0	59.8

Table 36 Complete data. Alternating EM algorithm. Backward elimination of parameters. Female and male variance components are constrained to their value of the more complex model.

It may be noted that the residual deviance, after having dropped out all the covariates is higher than the deviance exhibited in the Table 34 where the variance parameters are unconstrained.

Among the 36 tested interactions between the fixed effects, only one was significant according to the likelihood ratio test — that between age of the woman and the quality of the sperm after thaw (χ^2=8.6). This interaction was in the direction of decreasing benefit from "good sperm" with increasing age of the woman. However, since the number of tested interactions was large and the χ^2 was quite small , we have little evidence of any

real interaction between the fixed effects. We have also verified the absence of any significant variation of heterogeneity as a function of covariates by including interaction between variance components and fixed parameters in the *MLn* model for variance components estimation. In particular, there is no interaction between azoospermia and the variance at woman's level. Thus the variance of the probability of conception does not differ significantly between the two groups of women (sterile husband versus husband having sperm of poor quality).

The fixed effects include covariates at the cycle level, the woman level, the donation level and the donor level. The conclusions of previous studies are summarized by CECOS Fédération and Lansac, in Gray et al.(1993) in the following words:

"Pregnancy rates are highest when the cervix is dilated, there is abundant cervical mucus, and spinnbarkheit 10 [components of the Insler score]. A woman is more hypofertile if she (...) is over 35 years old, or has a husband with good semen. The most useful predictor of male fertility is spermatozoa mobility. (...) The most predictive variable [concerning male factors] was post-thaw mobility."

Concerning day of insemination, Schwartz *et al.* (1979) reported a large decrease of fecundability rates for inseminations taking place too early or too late, after the presumed day of ovulation. Our findings confirm these results. The results provided for standardized covariates are separately more easily interpretable as odds ratio : Age of the woman, (per 5 years) : 0.76; azoospermia of the husband : 1.18; Insler score, (per 1 unit) : 1.56; early insemination : 0.58; late insemination : 0.83; number of spermatozoa, (per 50 10^6 /ml) : 1.2; percentage of motile spermatozoa, (per 10 %) : 1.39 and the global index of quality, (per 1 unit) : 1.30. All the covariates quoted above, together with stimulation of ovulation by clomiphene citrate are significantly related to the success rate.

This last covariate could be partly a surrogate for dysovulation, this treatment being used in such cases. But clomiphene citrate is also known to be deleterious for the cervical mucus

and as such could cause of low success rate. The distinction between these two origins of low success rate when using clomiphene citrate is of interest. We can distinguish between these two sorts of effects by including the covariate "clomiphene at first cycle" as an indicator for low fertility. Additionally including the variable at the cycle level can be interpreted as a direct effect. This is an example of "compositional covariate". Table 37 shows the absence of reduction of deviance when adding the covariate "clomiphene citrate on the first cycle" (18.8 percent of the women). Thus, we do not have any argument to confirm the interest of this covariate as identifying less fertile women. Our observation suggests the existence of a negative effect of citrate clomiphene on the success rate.

Parameter	Estimate (s.e.)	
Intercept	-2.303	-2.303
Woman :		
Age (woman)	-0.107(0.036)	-0.106(0.036)
Azoospermia (husband)	0.079(0.037)	0.080(0.037)
Cloniphene citrate at first cycle	-0.036(0.044)	_
Cycle :		
Insler score	0.264(0.039)	0.263(0.039)
Early insemination	-0.136(0.038)	-0.136(0.038)
Late insemination	-0.082(0.034)	-0.082(0.034)
Cloniphene citrate	-0.086(0.042)	-0.103(0.036)
Donation :		
Sperm count	0.130(0.030)	0.130(0.030)
Sperm motility	0.179(0.034)	0.179(0.033)
Sperm quality	0.217(0.036)	0.217(0.036)
Heterogeneity :		
Between women	0.505(constrained)	0.505
Between donors	0.222(constrained)	0.222
-2 log likelihood (approximate)	6 876.6	6877.2

Table 37 Complete data. Alternating EM algorithm. With covariates.

It has been previously stated that "The success rate was considerably lower for women inseminated for the first pregnancy than for women seeking a second or a third pregnancy. This suggests that the former group includes some hypofertile women". This statement favors the existence of a heterogeneity of

basal fecundability between the women. In our datasets the residual variance between women is estimated at 0.5 and earlier analyses suggest it is probably rather larger. A much clearer interpretation can be obtained if we recall that the exponential of the random parameters (f and m) can be interpreted as (log) odds ratios between the basal fecundability of each woman (or donor) and the mean basal fecundability of the population of women (donors). These parameters being distributed normally, 26 percent of women (donors) have an odds ratio higher than e^{σ}. The size of the standard deviation of a random parameter can in this way be compared with a fixed effect parameter's. For women, the variance is 0.5 (standard deviation 0.7) this heterogeneity factor is three times larger than the greatest fixed effect (Insler score $\hat{\beta} = 0.264$). For donors the estimate of residual heterogeneity is smaller than among women (standard deviation: 0.47). This heterogeneity remains after having introduced the parameters describing the semen, before and after thaw.

Table 38 presents an analysis of the conception rate according to the three male characteristics ([i] Number of spermatozoa [ii] Motility of spermatozoa and [iii] Quality of sperm after thaw) introduced in several different ways. We should recall that studies of artificial insemination by donor represent an unusual research opportunity to study both male and female fertility simultaneously. In "normal" couples, these aspects are nearly totally confounded, while in these data the non-systematic allocation of the donor to recipient allows the effect to be differentiated.

The large number of cycles with insemination by the same donor allows a more detailed analysis of donor fertility than for fertility of the woman. Table 38 compares the variance component concerning the donors with or without inclusion of the three male covariates. The first model (I) does not include the male covariates. The second (II) introduces them as was stated in Table 35, directly at the donation level. Their inclusion decreases

significantly the deviance (χ^2= 35.6 for 3 df p<0.0001). The male residual heterogeneity decreases from 0.382 to 0.222 when these covariates are included.

Regression parameters	Estimates Model I	Model II	Model III
Donation level; male covariates			
Number of spermatozoa	-	0.130	-
Motility of spermatozoa	-	0.179	-
Quality of sperm after thaw	-	0.217	-
Donor level (means)			
Number of spermatozoa	-	-	0.179
Motility of spermatozoa	-	-	0.175
Quality of sperm after thaw	-	-	0.366
Donation level; male covariates (Centred on donor means)			
Number of spermatozoa	-	-	0.059
Motility of spermatozoa	-	-	0.110
Quality of sperm after thaw	-	-	0.099
Variance components			
Women	0.460	0.500	0.504
Donors	0.382	0.222	0.220
-2 loglikelihood (approximated)	6916.2	6880.6	6861.6

Table 38 Complete data. Male covariates introduced at donation level (Model II) or as donor compositional covariates (Model III). The female covariate effects are not presented being nearly unchanged. (identical in the first 2 decimal places)

Moreover Table 38 compares the model (II) with a model (III) in which the same three male covariates are no more included in their original form but as "compositional" covariates:

we include them in two parts (i) the mean for each donor, and (ii) the deviation of the measurement at each donation from this mean value. As an example, if we consider the quality of the sperm after thaw. The donor mean of this variable provides a global index of sperm quality from that donor, while the deviations measure whether sperm quality at any donation was good or bad for that donor. The table shows clearly a significant reduction of

the residual deviance (χ^2= 19 for 3 df p<0.001). Thus, this new way of introducing the variable carries more information. Moreover, the log-odds is originally 0.217 when introduced as a single covariate. When introduced at both levels, it becomes 0.366 at donor level and only 0.099 between donation. Analysis in this way clearly shows that the three covariates (the number of spermatozoa, their motility and the score of quality of the sperm after thaw) are more predictive when introduced at the donor level. In fact, the size of the estimated effect of the donor mean is three times as big as the residual effect of the deviation of each measurement from this mean.

More than one third of the heterogeneity between donors appears to be explained by these three covariates. Two explanations can be proposed : either their distribution among donors vary extremely and they act at the cycle level or more probably they are descriptors of the donors more than of each specific donation. Figure 13 illustrate the size of the overlapping between ranges of values observed in donations of each of th 279 donors.

The figures suggest that large values were observed for all donors whatever their average quality. This observation tends to reject the first explanation.

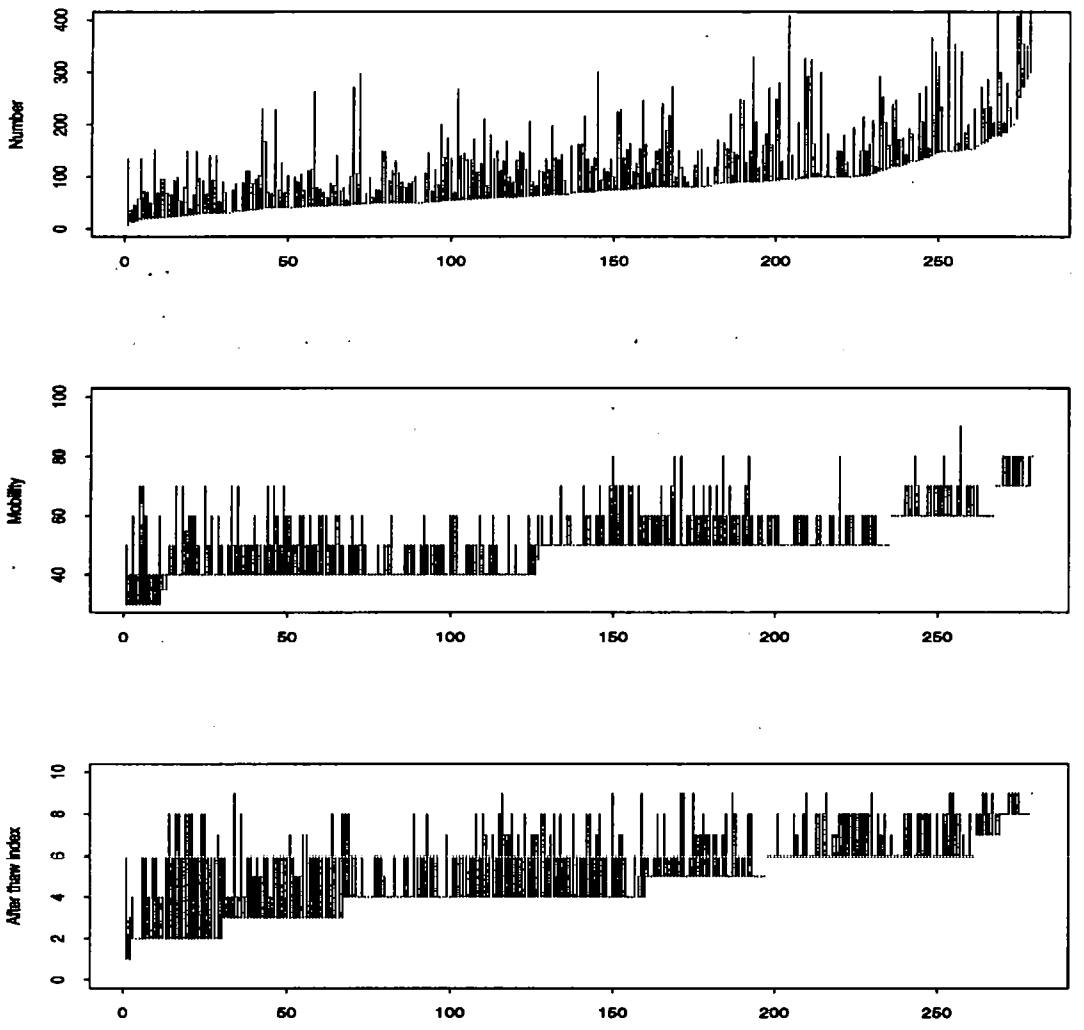


Figure 13 Range of values observed among donations of the 279 donors. The donors are sorted by the lowest value of the range.

Thus, these three male variables would seem to represent imperfect measurements of the global fecundability of the donor rather than indicators of the specific fecundability of one donation. This observation was made in the past concerning the segmentation of the embryo after in vitro fertilization (Lornage, Mathieu 1989): the percentage of successes was quite the same with sperm having good or bad indexes (number, motility etc...) if the

man had had, on average, good sperm in the past. On the other hand, if the man had past bad indices, good sperm used for the fertilisation failed to give good results. Finally, the very low variance (almost 0) obtained when we introduce donation as an intermediate level between donor and cycles seems to argue in the same direction; the quality of the donations of a same donor are very stable whatever the values of measured covariates. This stability of the quality of sperm between donations has been observed in other species (Goffaux 1978, Foote 1980). Our observation is not contradictory to the fact that all studies of repeat semen samples show a large variability within subject : this observed heterogeneity concerns the characteristics of the semen, not its fecundability; moreover in AID case, donations have been collected from a single donor over about a month's time and hence the heterogeneity may be reduced. A last analysis of this aspect shall be proposed in the last Chapter.

Recall that marginally the success rate of a donor increases as the number of previous use of his semen rises (Chapter 4). The unit-specific regression model used in this Chapter allows to test for this effect avoiding the selection bias due to any specific allocation made by the physician. The selection process made by the physician selecting for more donations sperms having a "good past history of success" shall not create any bias in this conditional analysis.

Table 39 shows the result of this analysis.

Parameter	Estimate (s.e.) A	Estimate (s.e.) B
Intercept	-2.328 (0.051)	-2.325 (0.051)
Cycle :		
Number of previous inseminations	-	0.034 (0.032)
Heterogeneity :		
Donors	0.382(constrained)	0.382 (0.059)
-2 log likelihood (approximated)	7 473.0	7 471.8

Table 39 Complete dataset. Donor at level II.

This table provides no argument for an effect of the number of previous inseminations on the success rates. The subject specific model does not support the hypothesis of an improvement effect of time on the frozen sperm. We conlude that the effect we observed Chapter 4 was a consequence of a selection bias.

3. Conclusion

The data on Intra uterine insemination with donor's sperm has a complex structure due to the overlapping of two hierarchies, one concerning female (recipient) factors (cycles within attempts within women), and the other concerning male factors (inseminations within donations within donors). A crossed multilevel model, taking into account these aspects permits one to improve the estimation of the fixed effects and provides insights into the influence of them at each level (cycle level, woman level and donor level). The penalized likelihood method applied successively to each of the two structures (male and female) after introduction of the estimation of the opposite structure as offset permits efficient computation in multilevel modelling software. Unfortunately we cannot claim general applicability for this computational method as its justification depends on near orthogonality of the crossed classifications. Our experience of stratification by year of first treatment is reassuring and suggests that modest non-orthogonality may be tolerated. More experience of this will be necessary before a general recommendations can be made. In the meantime we would stress that, short of Markov chain Monte Carlo methods, no other algorithm was available to us which could fit a crossed random effect model of this size. We would also note that the circumstances in which this method will not work, that is strongly non-orthogonal designs, are precisely those in which the method of Rasbash and Goldstein (1994) can be applied. At the very least our method provides, very rapidly, an approximate solutions which could be refined, for example, by using MCMC methods. This will be investigated in the next Chapter.

Chapter 8 A Gibbs sampling approach

In previous chapters we fitted the logistic Gaussian mixed model to our data set using approximate inference methods. This approximation was necessary to overcome the difficulty due to the need for high order numerical integration. Recent Bayesian procedures solve these problems by taking repeated *samples from the posterior distribution* using Gibbs sampling techniques; effectively Monte Carlo techniques are used to carry out the integration. In this Chapter, this approach will be presented and applied to our data set.

1. Bayesian formulation

In a Bayesian approach to analysing the logistic Gaussian mixed model, the parameters, β and θ are random variables. Recall that the results of ovulatory cycles in terms of success or failure are modelled as follow

$$y \sim Bernouilli(\lambda)$$

$$g(\lambda) = \log\frac{\lambda}{1-\lambda} = \eta$$

$$\eta = X\beta + Zb$$

$$b \sim N(0, D(\theta))$$

In the Bayesian setting, the important distinction is that between observable (Y, X and Z) and unobservable quantities (β, $D(\theta)$ and b). No further classifications are necessary. In particular, both fixed effects and random effects are handled within the same general

framework. The Bayesian model specification is completed with prior distributions. We choose the following prior distributions

$$\beta \sim Normal(\beta_0, B)$$

$$\theta = \frac{1}{\delta} \text{ and } \delta \sim Gamma(\nu, \tau)$$

The parameters of the distribution of random effects are commonly termed "hyperparameters":

β are parameters and θ is a "hyperparameter". Thus (β_0, B) and (ν, τ) are respectively called priors and hyperpriors.

Unobservables quantities are estimated using the distribution obtained by conditioning on all observables, and integrating over all other unobservables. Let $p(\beta, D)$ represent the joint prior distribution for β and $D(\theta)$. The first objective of our analysis is to derive the posterior distribution, given by

$$f(\beta, D|y) = \frac{\prod_{i=1}^{I} \int f(y_i|b_i, \beta) g(b_i|D) p(\beta, D) \partial b_i}{\prod_{i=1}^{I} \int f(y_i|b_i, \beta) g(b_i|D) p(\beta, D) \partial b_i \partial \beta \partial D}$$

i.e. conditioning on the data (y_i) and integrating over other unobservables (b_i), *i=1,...,I*, where *I* is the number of clusters (i.e. women). The second concerns the random parameters

$$f(b_i|y) = \frac{\int f(y_i|b_i, \beta) g(b_i|D) p(\beta, D) \partial \beta \partial D}{\int f(y_i|b_i, \beta) g(b_i|D) p(\beta, D) \partial b_i \partial \beta \partial D}$$

i.e. conditioning on the data (y_i) and integrating over other unobservables (β and D). Numerical evaluation of either $f(\beta, D|y)$ or $f(b_i|y)$ is typically intractable. In the logistic Gaussian mixed model we seek the joint distribution $[\beta, D, b|y]$ and its marginals $[\beta, D|y]$

and $[b|y]$. These posteriors are not directly available but may be obtained from conditionals as will be shown now.

2. Gibbs sampling

The Gibbs sampler is a method for obtaining the marginals from a set of full conditionals This procedure avoids the need for numerical integration by taking repeated sampling from the posterior distribution.

Stochastic substitution

For clarity, let us limit the presentation to the case of 2 elements (rather than three), the generalization to more than two elements being straightforward (see Clayton, 1991). To draw a sample from the unavailable $[u,v]$, if $[v]$ is available, sample V from $[v]$, then sample $U\sim[u|v=V]$, thus (U,V) is a sample from $[u,v]$. If $[v]$ is not available, argument by analogy with relaxation methods for the solution of linear simultaneous equations suggests a *stochastic substitution* method : starting from (U^0,V^0) we generate a sequence (U^{it},V^{it}), *it=1,2,...* sampling $U^{it}\sim[u|v=V^{it-1}]$ and $V^{it}\sim[v|u=U^{it}]$. Using theorems reviewed by Gelfand and Smith (1990) it may be shown under rather weak conditions that the sequence, which is Markovian, converges to an equilibrium distribution which is the joint distribution $[u,v]$

For more than two sets of variables, the basic idea of stochastic substitution may be extended in various ways according to the availability of conditional distributions.

The most attractive is the Gibbs sampler algorithm of Geman and Geman (1984) which was initially constructed to generate random samples from the Gibbs distributions — a very large class of graphical models.

The graphical model

A partial ordering of subelements of our hierarchical model is obtained using a Directed Acyclic Graph (DAG, Figure 14). Conditionally independent submodels are then identified as illustrated on a conditional independent graph (Figure 15). The joint distribution of the system is proportional to the product of each independent part. Such a distribution is a Gibbs distribution.

Figure 14 describes the two level hierarchical model (women and cycles) through a directed acyclic graph connecting the hyperparameters, the parameters and the observed data. β_0 represents the intercept, i.e. the basal fecundability supposed to be constant over the short period of time covered by our study. β_k represents parameters corresponding to fixed covariates whatever be the level (women or cycles). Finally, f represents random parameters and θ_f the hyperparameter. It appears clearly that the respective position of parameters and hyperparameters are not identical. A *partition* into conditionally independent submodels follows naturally.

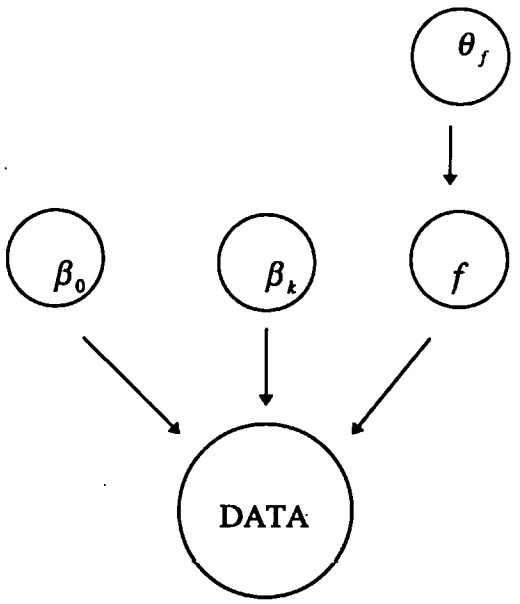


Figure 14 Female two level hierarchical model

Directed acyclic graph

In Figure 14 the edges are directed and it is not possible just by following the direction of edges, to return to a node after leaving it. Thus, there is a *partial ordering* of the nodes : a natural Markov-type assumption on this partial ordering leads to a "directed Markov assumption" (Lauritzen, 1990), which simply states that the joint distribution of all the model parameters and data is given by the product of all the submodels exhibited on the conditional independence graph (Figure 15).

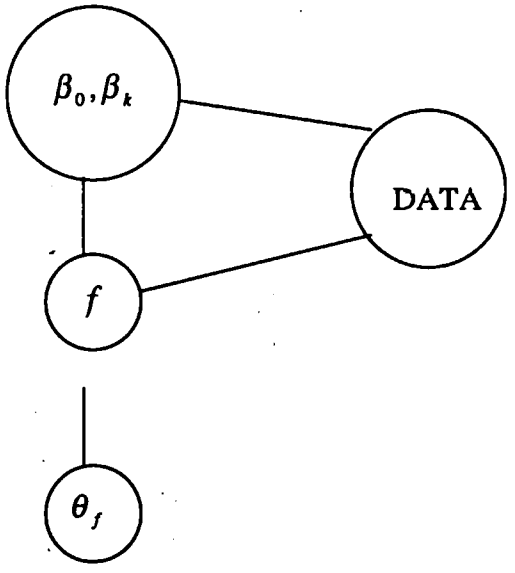


Figure 15 Conditional independence graph

The conditional independence graph is an undirected graph constructed so that if two variables, U and V, are connected only via a third variable W, then U and V are conditionally independent given W; if U and V both have directed links to W in the directed graph, then U and V must be joined in the (undirected) conditional independence graph. These independences lead to considerable simplifications as will be exemplified below. The joint distribution, a product of each independent part described above is a Gibbs distribution. The Gibbs distribution is defined on the two "*cliques*" — set of nodes in which all pairs are connected.

Gibbs sampling

The Gibbs sampler (Gelfand and Smith, 1990; Casella and George, 1992) is a Monte Carlo method for estimating the desired posterior distributions. The model is first analysed adequately as shown above (Figure 14 and Figure 15). Given arbitrary starting values, the Gibbs sampler algorithm visits each node of the conditional independent graph and generates a value from the conditional distribution of the corresponding random variable

given the current values of all its neighbours. Geman and Geman (1984) showed that, under conditions similar to those for simple stochastic substitution, the algorithm converges to a sample from the joint distribution. Gibbs sampling succeeds because it reduces the problem of dealing simultaneously with a large number of intricately related unknown parameters and missing data into a much simpler problem of dealing with one unknown quantity at a time.

Example of an application for the logistic Gaussian mixed model

As previously said we seek the joint distribution $[\beta,b,D|y]$ which can be obtained by sampling from the "full conditional distributions" $[\beta|D^{[it]},b^{[it]},y]$, $[D|\beta^{[it]},b^{[it]},y]$ and $[b|\beta^{[it]},D^{[it]},y]$. These conditional distributions simplify because of the conditional independence argument :

$[\beta|D^{[it]},b^{[it]},y]$ becomes $[\beta|b^{[it]},y]$, and $[D|\beta^{[it]},b^{[it]},y]$ becomes $[D|b^{[it]}]$.

We can then specify the simulation method for each node (Zeger and Karim, 1991), as follows.

$[\beta|b^{[it]},y]$

Given the $b^{[it]}$'s, the random effects model reduces to a GLM with offset $z^T b^{[it]}$ for each observation. Assuming a flat prior for β, the posterior is proportional to the likelihood. To sample from this likelihood we can choose between two options :

> either approximate it with a quadratic function in the neighborhood of $\beta^{[it]}$
> or, a second solution, prefered in smaller samples, is to obtain a sample from the likelihood itself using a rejection/acceptance algorithm by sampling the distribution

$$cN\left(\hat{\beta}^{[it]}, 2 * V_{\beta}^{[it]}\right)$$

where c is selected so that the mode of this function and of the likelihood are equal and $V_{\beta}^{[it]}$ is the inverse of the information matrix

$$\left[D \middle| b^{[it]}\right]$$

We have assumed that the b_i's are independent Gaussian$(0,D)$ random variables.

The standard non-informative prior for D is $|D|^{-\frac{q+1}{2}}$ where q is the number of random parameters. Then the posterior distribution of D^{-1} follows a Wishart distribution with parameters

$$S^{[it]} = \sum b_i^{[it]} b_i^{[it]T} \text{ and } (I\text{-}q\text{+}1) \text{ df.}$$

$$\left[b \middle| \beta^{[it]}, D^{[it]}, y\right]$$

is obtained using again a rejection/acceptance algorithm with the function

$$cN(A, 2 * B)$$

A and B being respectively the Bayes estimate of b and its variance $\left(Z^T V Z + D^{-1}\right)^{-1}$.

Sampling successively from each of these three conditional distributions provides samples from the joint posterior distribution of the unknown quantities. Empirical summary statistics can be formed from these samples and used to draw inferences about their true values. Practical aspects concerning the convergence matters will be discussed below.

3. BUGS

BUGS (Gilks, Thomas and Spiegelhalter, 1994; Spiegelhalter et al, 1995; Best et al, 1996) is the most well known software for "Bayesian inference Using Gibbs Sampling". BUGS Version 0.50 [i] *decomposes* the Directed Acyclic Graph -*DAG*- model into its components, the nodes, [ii] identifies the nature of parents and children of each *node* and thus the structure of their full conditional distributions, [iii]choose the *best method of sampling* from full conditionals and provides the samples. Its presentation introduces a generalized approach of Gibbs sampling.

Decomposition of the Directed Acyclic Graph -DAG- model into its components, the nodes

The DAG model represents, as previously said, parameters, data (including missing) and constants as nodes in a directed graph. These nodes are reached by arrows which origines rom other nodes having a direct influence on their values (parents). Formally, the model represents the assumption that, given its parent nodes, each node v is independent of all other nodes in the graph except "descendants" of v (see BUGS Version 0.50 manual, Spiegelhalter et al, 1995). Nodes are of three types : [i] Constants, fixed by the design of the study, for AID data, the number of women, number of cycles,etc... [ii] Stochastic nodes, variables that are given a distribution. They may be observed (the data) or unobserved (parameters,missing data or unobserved due to censoring). [iii] Deterministic nodes, which are logical functions of other nodes.

Structure of the full conditional distribution of each node

(Best et al, 1996)

The *full joint distribution* of all the quantities V in the model has a simple factorization in terms of the conditional distribution $p(v|parents[v])$ of each node v given its parents (i.e. those nodes on whom v directly depends), i.e.

$$p(V) = \prod_{v \in V} p(v|parents[v])$$

Thus we only need to provide the parent-child distributions to specify the model fully : the crucial idea behind BUGS is that this is all we need to provide for being ble to carry out the whole analysis. The Gibbs sampling algorithm successively samples from the conditional distribution of each node given all the others. For any node v the *full conditional distribution* is the product of all the terms in V containing v, and thus, has the form

$$p(v|parents[v]) \times \prod_{v \in parents[w]} p(w|parents[w])$$

or in other words "prior component * likelihood components arising from each child of v". "The full conditional distribution for any node depends only on the values of its parents, children and co-parents, where 'co-parents' are other parents of the children of v." (Best et al, 1996).

Choice of the best method of sampling from full conditionals

BUGS contains a small expert system for deciding the best method of sampling from full conditionals. The choices are, in decreasing order of preference

(1) Conjugates : BUGS identifies conjugacy and takes benefit of it

(2) Adaptative Rejection Sampling (Gilks and Wild, 1992) for log concave densities, i.e. $\frac{\partial^2 \log f(x)}{\partial x^2} \leq 0$.

(3) Other methods in case of non-log-concavity. BUGS manual proposes a solution in case of non log-concave distribution. It is possible to discretise it into a categorical variable which can be sampled by enumeration.

In further Sections we will present some results provided by BUGS for AID data.

4. Application to AID data set

In this Section we present a description of the specification of three models and the results of their fitting using BUGS. We shall present successively a model including a random effect for women, a model including a random effect for men, and finally a model including there two random effects and covariates.

Variance of the probability of conception among the women

Using the results of the ovulatory cycles of the first attempt, as it was done in the first Chapters of this dissertation, we analyse the variance of the probability of conception among the women. Recall that this variance may be partly related to the heterogeneity between the women and partly to the heterogeneity between the donors whose sperm has been used for the inseminations. In previous Chapters the variance has been estimated fitting various models to the data : a geometric model with overdispersion based on the conception rates of the first two cycles (Chapter 3), a beta-geometric model describing the delays until conception as a mixture of geometric (Chapter 3), a gamma-geometric model allowing to include fixed effects properly (Chapter 4). In this Chapter we fit a logistic Gaussian mixed model, mainly in order to be able to compare the results of these various approaches.

BUGS code for logistic Gaussian random intercept model :

Note that for computational reasons, the normal distribution is parameterized in terms of mean and precision, where precision is the inverse variance.

The symbol # denotes a comment.

```
const
   F=1901,                                  # Number of women
   CYCLE=9740;                              # Number of cycles
var
   y[CYCLE],                                # y, success or failure
   lambda[CYCLE],                           # λ, fitted value
   woman[CYCLE],                            # woman's identication
   randpar[F],                              # f, random effect
   deltaf,                                  # δf, precision
   intercept,                               # β0, intercept
   thetaf;                                  # θf, the variance par.
data in "firstat.dat";                      # the data set
inits in "firstat.in";                      # Initial values
{
for(k in 1:CYCLE) {
   logit(lambda[k]) <-
        intercept+randpar[woman[k]];        #  logit(λ) = Xβ + Ff
   y[k] ~ dbern(lambda[k]);                 #  y ~ Bernoulli(λ)
}
for(i in 1:F) {
   randpar[i] ~ dnorm(0.0, deltaf);         #  f ~ Normal(0, 1/δf)
}
intercept~ dnorm(0.0, 1.0E-6);              #  Priors β ~ Normal(β0, B)
deltaf ~ dgamma(1.0E-3,1.0E-3);             #  δ ~ Gamma(ν, τ)
thetaf <- 1/deltaf;                         #  and θf = 1/δf
}
update(1000)                                # Burn-in iterations
monitor(intercept)                          # To store all sampled
monitor(thetaf)                             # values for the nodes
update(3000)                                # Carry out 3000
                                            # iterations
```

The results and those obtained in the previous Chapters are showed Table 40 for comparison. Four methods of estimation are used : a moment method, a maximum likelihood method, an approximate likelihood (*PQL*) and the Gibbs sampling.

Model	Estimation Method	Variance on logit scale
Overdispersed geometric model	Moment method*	1.564
Beta geometric model	Likelihood*	0.740
Logit Gaussian model	Approximated likelihood(*PQL*)	0.431
Logit Gaussian model	Gibbs sampling	0.895

* estimated on the hazard scale and approximated on the logit scale using the delta method and thus slightly biased

Table 53 Two level models Women at level II First attempt

The ML estimate of the variance on the complementary log-log scale, obtained fitting a gamma-geometric model to the delays until conception is 1.02. The higher variance (1.564) obtained using the marginal rates of the first two cycles may be interpreted in at least two ways: either sampling variation or the existence of a small group of highly fertile women conceiving during the first cycle. So far we do not have any added information to choose between these two hypotheses. Likelihood and Gibbs sampling methods provide similar results and *PQL* lower estimations. This can be interpreted as a downward bias of *PQL* estimates.

Our Gibbs sampling results may also be slightly biased downward since we had a delayed convergence (Figure 16) and some difficulty to obtain stability (Figure 17).

In these figures we present the result of the series of iterations, with both a kernel density (on the left) and the complete series of values (on the right). Figure 17 represent the results of the same sampling but after having excluded the burn in iteration. A bimodal density of the sampled values reflect the lack of stability of the sampling process. Nevertheless this bimodal distribution could also be the true distribution !

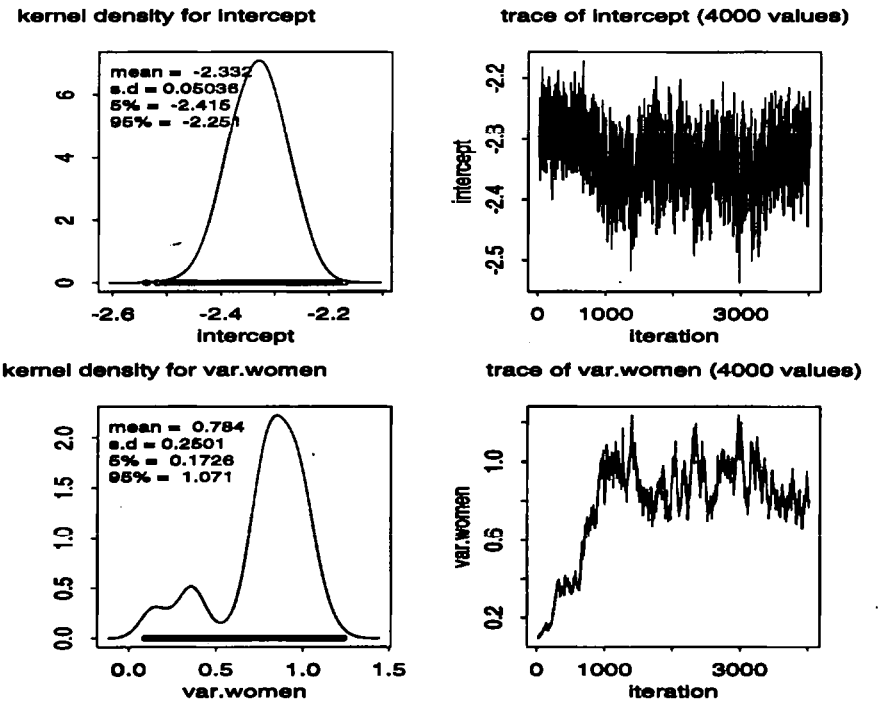


Figure 16 Logistic Gaussian mixed model. Female hierarchy. First attempt. Gibbs sampling

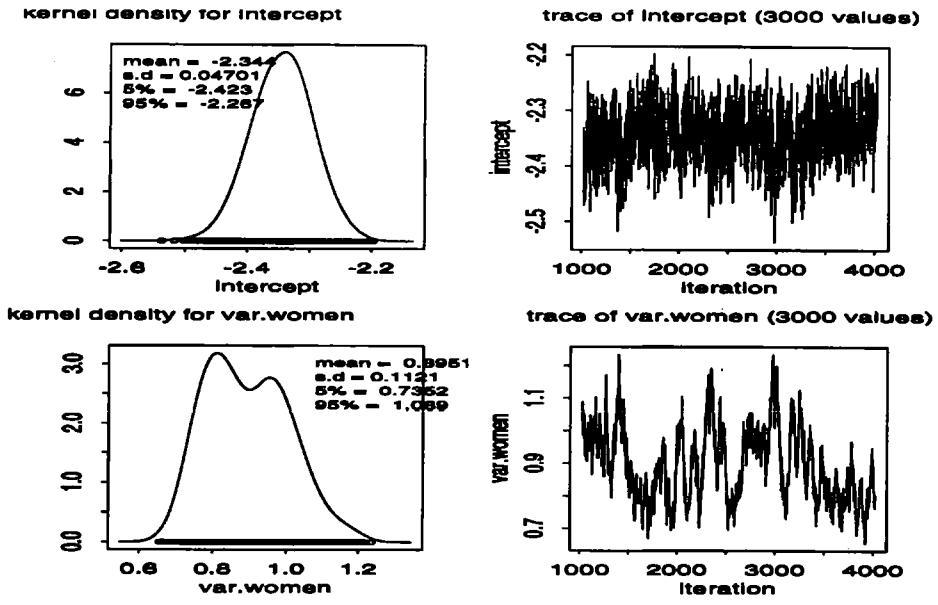


Figure 17 Logistic Gaussian mixed model. Female hierarchy. First attempt. Gibbs sampling

Variance of the probability of conception among the donors

After having studied the variance of the probability of conception between women, we estimate the variance of the probability of success among donors. As it is the case for the women the variance is not reflecting only the heterogeneity between the donors. Nevertheless the number of recipients per donor being high the added variance due to them is possibly low.

We present the BUGS code used to fit a beta binomial model. r represents the number of success per donor, λ is the probability of success, described as beta distributed.

BUGS code for the beta binomial model

```
const
D =279;                                  # Number of donors

var
   r[D],                                 # r,number of successes
   m[D],                                 # m,number of trials
   lambda[D],                            # λ, fitted value
   eta,                                  # gamma parameters
   tau;
data in "betabin.dat";                   # The data set
inits in "betabin.in";                   # Initial values
{
for(k in 1:D)
   {r[k] ~ dbin(lambda[k],m[k]);         #  Bernouilli
}
for(i in 1: D)
   {lambda[i] ~ dbeta(eta, tau);         #  Beta distribution
}
eta~ dunif(0,1000);                      #  Flat priors
tau ~ dunif(0,1000);
}
update(500)                              #  Burn-in iterations
Monitor(eta)                             #  Store sampled values
Monitor(tau)
update(3500)                             #  3500 iterations
```

Note that the convergence was obtained rapidly and that the stability was sufficient as shown Figure 18.

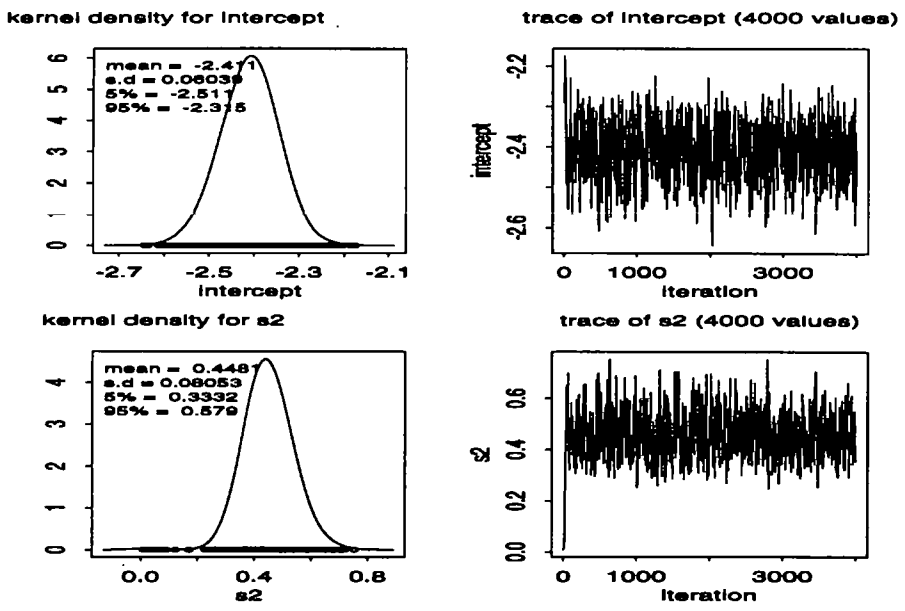


Figure 18 Beta binomial model. Male hierarchy. Complete data. Gibbs sampling

Table 41 shows the results concerning the the variance of the probability of success among donors, estimated on a same scale (logit scale), fitting different models with various inference methods.

Model	Estimation Method	Variance on logit scale
Overdispersed binomial model	Williams method*	0.426
Beta binomial model	Likelihood*	0.448
Beta binomial model	Gibbs sampling*	0.423
Logit Gaussian model	PQL	0.387
Logit Gaussian model	Gibbs sampling	0.448

* estimated on the hazard scale and approximated on the logit scale using the delta method

Table 41 Two level models Donors at level II Complete data set

The various methods provide very similar results; in particular, the result obtained using *PQL* is only slightly lower than others, in contrast to the heavy downward bias observed in the analysis of the woman random effect.

Logistic Gaussian model for crossed hierarchies : female and male

We now present the crossed random multilevel logistic model, taking account of the crossed hierarchical structure due to the coexistence of female factors and male factors. This model was applied to our data in the previous Chapter using an approximate inference method based on the *PQL* approach. We present below the specification of the model in BUGS and the results.

BUGS code :

```
const
   F=1901,                                  # Number of women
   M=279,                                   # Number of donors
   CYCLE=12100;                             # Number of cycles
var
   y[CYCLE],                                # y, success or failure
   lambda[CYCLE],                           # λ, fitted value
   agewcent[CYCLE],jplus1or[CYCLE],         # covariates
   clomiphen[CYCLE],mobcent[CYCLE],         #
   numcent[CYCLE],inscent[CYCLE],           #
   jm3orles[CYCLE],deccent[CYCLE],          #
   azoo[CYCLE],                             #
   donor[CYCLE],                            # donor's identication
   woman[CYCLE],                            # woman's identication
   randparwo[[F],                           # f, random effect
   randpardo[M],                            # m, random effect
   deltaf,                                  # δf, precision
   deltam,                                  # δm, precision
   intercept,                               # b0, intercept
   betaagewcent,betajplus1or,               # other fixed effects
   betaclomiphen,betamobcent,               #
   betainscent,betanumcent,                 #
   betajm3orles,betadeccent,betaazoo,       #
   thetaf,thetam;                           # θfand θm the variances
data in "wodocov.dat";                      # the data set
inits in "wodocov.in";                      # Initial values
{
for(k in 1:CYCLE) {
   logit(lambda[k]) <-                      #     logit(λ) = Xβ + Zb
       intercept+                           #
       betaagewcent*agewcent[k]+            #
       betajplus1or*jplus1or[k]+            #
       betaclomiphen*clomiphen[k]+          #
       betamobcent*mobcent[k]+              #
       betainscent*inscent[k]+              #
       betajm3orles*jm3orles[k]+            #
       betadeccent*deccent[k]+              #
       betanumcent*numcent[k]+              #
       betaazoo*azoo[k]+                    #
       randparwo[woman[k]] +randpardo[donor[k]];
   y[k] ~ dbern(lambda[k]);                 #     y~ Bernoulli(λ)
}
for(i in 1:F) {
```

```
randparwo[i] ~ dnorm(0.0, deltaf);          # f ~ Normal(0, D(θ_f))
}
for(j in 1:M) {
    randpardo[j] ~ dnorm(0.0, deltam);      # m ~ Normal(0, D(θ_m))
}
intercept~ dnorm(0.0, 1.0E-6);              # Priors β ~ Normal(β_0, B)
betaagewcent~ dnorm(0.0, 1.0E-6);           #
betajplus1or~ dnorm(0.0, 1.0E-6);           #
betaclomiphen~ dnorm(0.0, 1.0E-6);          #
betamobcent~ dnorm(0.0, 1.0E-6);            #
betanumcent~ dnorm(0.0, 1.0E-6);            #
betainscent~ dnorm(0.0, 1.0E-6);            #
betajm3orles~ dnorm(0.0, 1.0E-6);           #
betadeccent~ dnorm(0.0, 1.0E-6);            #
betaazoo~ dnorm(0.0, 1.0E-6);               #
deltaf ~ dgamma(1.0E-3,1.0E-3);             # δ ~ Gamma(ν, τ)
deltam ~ dgamma(1.0E-3,1.0E-3);             #
thetaf <- 1/deltaf;                         #    and θ = 1/δ
thetam <- 1/deltam;                         #
}
```

It must be pointed out that this crossed model is close to the limit of feasability with current equipment: one hour was needed on a Pentium PC 75 for compilation and 1000 updates took two days !

After these burn-in iterations, 3000 iterations was used to obtain the following results.

Parameter	Estimate (s.e.)
Intercept	-2.609(0.062)
Woman :	
Age (woman)	-0.136(0.043)
Azoospermia (husband)	0.096(0.043)
Cycle :	
Insler score	0.297(0.044)
Early insemination	-0.155(0.040)
Late insemination	-0.089(0.035)
Clomiphene citrate	-0.112(0.039)
Donation :	
Sperm count	0.140(0.047)
Sperm mobility	0.203(0.048)
Sperm quality	0.237(0.048)
Heterogeneity :	
Between women	0.993(0.146)
Between donors	0.335(0.062)

Table 42 Complete data.

This Table has to be compared to the Table 35, which exhibits the results obtained fitting the same model using *PQL*. All the fixed parameters were closer to *0* with *PQL*, and the variance between women and between donors was also respectively 50% and 30% lower than results presented Table 42. These results confirm the well known downward bias of *PQL*.

At the end of this seventh Chapter, dealing with Gibbs sampling, we conclude first with two statements:

"From now on we can compare our data with the model that we actually want to use rather than with a model which has some mathematically convenient form".

(Clifford, discussions of Gilks *et al.*, 1993)

Indeed, Gibbs sampling offers a solution to estimate fixed and random parameters in our hierarchical model whatever its complexity (more than two levels, crossed hierarchies,...) Since this method of estimation is based on empirical means and standard deviation of processes which are supposed to have converged to a limiting distribution it is very important to be able to check the convergence and other characteristics of the sample obtained. Therefore this method should be used with some caution (Spiegelhalter et al., 1994).

It is necessary to check very carefully the adequacy of the posterior distribution sampling process before calculating resulting estimators. Moreover, there is a need for further research to speed up this sampling process and for further advances in computing technology. Nevertheless, this method provides an additional and promising tool for the analysis of data on conception delays.

Chapter 9 Further topics, discussion, conclusion

In this dissertation we have emphasized some important aspects of the statistical treatment of AID data, but many related questions of interest were not treated or just outlined. This Chapter presents briefly some of them, selected for their practical implications or because they provide potential areas for further developments.

We have stressed that, using mixed models to describe our data, inference provides estimates of fixed effects, variance components and random effects. So far the calculation of empirical Bayes estimates of the latter was introduced only as a useful aid in the algorithm of estimation of the components of the variance but their properties as an estimate of the performance of a woman or of a donor were not considered; this will be discussed in the present Chapter. Then some open problems shall be outlined, including propositions made by some authors to improved methods, which rely on various approximations of the likelihood.

The advantage of random effects models in the statistical analysis of human fecundability data is emphasized. A rapid overview of potential benefit of their use in that field closes the dissertation.

1. Distribution of the random effects

We shall discuss successively the presence of some sterile women among the recipient and more generally the distribution of the random effects among the women and among the donors. In earlier Chapters we assumed that neither any woman nor any donor are totally

sterile. Indeed both, recipient and donors are accepted for intra uterine insemination only if there is a strong belief in their fertility. But a sizeable proportion of women being unable to conceive is known to exist in any population (Leridon, 1977), and this status may be unknown at the start of an attempt for fertilization.

Sterile recipients

All evidence for a sterility of the woman would be a contra-indication for AID : if any, the percentage of sterile women is probably very low. A way of estimating the percentage of steriles was proposed by Leridon from a comparison of rate of increase of the family at different parities. The mover-stayer model supposes that some women whose fecundability is effectively 0 are present in the dataset.

For such applications, the beta distribution (for example) could be considered to be contaminated by a distribution degenerated at 0 (Maruani and Schwartz, 1983; Weinberg and Gladen, 1986).

The distribution is then a mixture of a continuous function and a Dirac measure δ^0, concentrated at 0 : Under non-informative censoring the marginal hazards are a mixture of a proportion of sterile women and of women being progressively selected, the more fertile conceiving earlier. Denoting s the proportion of sterile women at the beginning of the study, $f(\lambda)$, the distribution of the probability of conception among the women is

$$f(\lambda) = s\delta^0 + (1-s)\frac{\lambda^{\nu-1}(1-\lambda)^{\tau-1}}{B(\nu,\tau)}$$

where B() denotes the beta distribution. For a woman of fecundability λ, under a geometric model, the probability of conception at cycle t is $\lambda(1-\lambda)^{t-1}$. By integration we obtain the distribution of the delay of conception, T say, among the women

$$\Pr(T=t) = 0.s + (1-s)\int \frac{\lambda^{\nu-1}(1-\lambda)^{\tau+t-2}}{B(\nu,\tau)}\partial\lambda$$

$$= (1-s)\frac{\nu}{\nu+\tau}, \text{ for } t=1$$

$$= (1-s)\frac{\nu}{\nu+\tau}\prod_{u=1}^{t-1}\frac{\tau+u-1}{\nu+\tau+u}, \text{ for } t>1$$

and the probability of no conception up to cycle t

$$\Pr(T>t) = s + (1-s)\prod_{u=1}^{t}\frac{\tau+u-1}{\nu+\tau+u-1}.$$

Note that these expressions are not exactly identical to the published expressions (Maruani and Schwartz, 1983). Indeed Weinberg (1986) has pointed out some typographical errors in this paper.

If $n(u)$ and $c(u)$ denote respectively the number of women conceiving and being censored at the tth cycle, the log-likelihood may be written

$$l = \sum_{u=1}^{12}\left\{n(u)\log[\Pr(T=u)] + c(u)\log[\Pr(T>u)]\right\}$$

We fit this model to the results of the first attempt on our dataset, using S-Plus

S-Plus Code

```
# nu and tau are the parameters of the beta distribution
# ss denotes s, the percentage of sterile women in the population
# g.maruani calculate Pr(T=t|ν,τ,s) and G.maruani Pr(T>t|ν,τ,s)
g.maruani <- function(tr, log.nu, log.tau,logit.ss)
{
   nu <- exp(log.nu)
   tau <- exp(log.tau)
   ss<-(1+exp(-logit.ss))^-1
   un.tr<-1:12
   comp<-((tau+un.tr-1)/(nu+tau+un.tr))
   prod.comp<-cumprod(comp)
   c((1-ss)*nu/(nu+tau),((1-ss)*(nu/(nu+tau))*prod.comp[(tr-1)]))
}
G.maruani<- function(tr, log.nu, log.tau,logit.ss) {
   un.tr<-1:12
   nu <- exp(log.nu)
   tau <- exp(log.tau)
   comp<-(tau+un.tr-1)/(nu+tau+un.tr-1)
   prod.comp<-cumprod(comp)
   ss<-(1+exp(-logit.ss))^-1
```

```
   ss+((1-ss)*prod.comp[tr])
}
lnn <- log(2)
lnt <- log(4)
inits<--2
bgm <- ms(~ -((event*log(g.maruani(tr, log.nu, log.tau,logit.ss)))
   +(censored*log(G.maruani(tr, log.nu, log.tau,logit.ss)))),
   start = list(log.nu= lnn, log.tau= lnt,logit.ss=inits))
```

The ML estimate of the percentage of sterile women is $\hat{s} = 1.7e-007$ i.e. effectively zero, and the mean and variance of the probability of conception among the women are respectively 0.0946 and 0.0056 (0.767 on the logit scale).

Note that the results are slightly different from those obtained fitting the marginal hazards with the same model (mean and variance of λ are respectively 0.12 and 0.008). We have actually underestimated the success rate in the present analysis because we have included attempts of women which were left truncated and considered them as complete observations.

Heckman *et al* (1990a and 1990b) presents nonparametric methods for testing the hypothesis of the existence of a mover-stayer model . We have not applied this approach to our dataset.

Non-parametric model for the distribution of the random effects

Non-parametric modelling of the distribution of the random effects is an attractive approach in that it allows for a bimodal (plurimodal) distribution. Doing that modelling we do not test for the presence of sterile women but more generally we describe the distribution of the random effects.

Let $(Y_1,...,Y_n)$ represent the binary response to a total of n cycles of insemination, where Y_i=1 if the ith cycle of insemination ends with a pregnancy, and Y_i=0 if not.The result of each insemination, y_i , are independent Bernoulli random variables with probability $\lambda_i = pr\{Y_i = 1\}$ satisfying

$$\text{logit}(\lambda) = X\beta + Zb$$

Let us note $f(y|\beta,b)$ the logistic distribution. the In the previous Chapters the random effects, b, was described as drawn from a Normal density distribution. Now this distribution, $H(b)$, is described as having K masses located at points α_k (k=1,...K) with probability mass p_k.

Recently Aitkin (1996) presented a general maximum likelihood analysis of overdispersion in generalized linear models. The authors stress that their aim is not to estimate the distribution — noting that the non parametric ML estimate of this may be very poor — but to avoid possibly misleading inferences from an inappropriate and unverifiable model assumption. Nevertheless we apply this method to our dataset, the approach providing complementary information on the diversity of probability of conception among the women and among the donors. The masses and mass-points are treated as unknown parameters. The number K of mass-points is also unknown but is treated as fixed, and sequentially increased until the likelihood is maximized. The likelihood is then written

$$L(\beta,\alpha_1,\ldots,\alpha_K,p_1,\ldots,p_K|y) = \prod_i^n \sum_{k=1}^K p_k f(y|\beta,\alpha_k)$$

where α_k and p_k are respectively the mass-points and the masses. The linear predictor in the kth mixture component is

$$\eta_k = X\beta + \alpha_k$$

Thus α_k functions as an intercept parameter for the kth component : it can immediately be estimated by including a 'component factor' in the model with K levels instead of the variable α_k .

$$\frac{\partial l}{\partial \beta} = \sum_i \frac{\sum_k p_k f_{ik} \frac{\partial \log f_{ik}}{\partial \beta}}{\sum_k p_k f_{ik}} = \sum_i \sum_k w_{ik} s_{ik}(\beta)$$

where w_{ik} is the posterior probability that observation y_i, comes from component k :

$$w_{ik} = \frac{p_k f_{ik}}{\sum_l p_l f_{il}}$$

and $s_{ik}(\beta)$ is the β-component of the score (the log-likelihood derivative with respect to β) for the observation i in component k. Equating the score to zero gives likelihood equations which are simple weighted sums of those for an ordinary GLM with weights w_{ik} . Alternately solving these equations for given weights and updating these weights from the current parameter estimates, is an EM algorithm. In each M-step the estimate of p_k is obtained from the weights :

$$\hat{p}_k = \sum_i \frac{w_{ik}}{n}$$

It may be shown that the transformation defined improves the likelihood at each iteration, which suggests iterating to convergence to obtain the maximum likelihood estimate. Model comparisons are carried out via the likelihood ratio test using differences of deviances. We apply this model and algorithm to our dataset using the GLIM4 implementation provided by Aitkin *et al* (1995).

A logistic random intercept model is fitted to the data for first attempts. The results are summarized in Figure 19 and Table 42

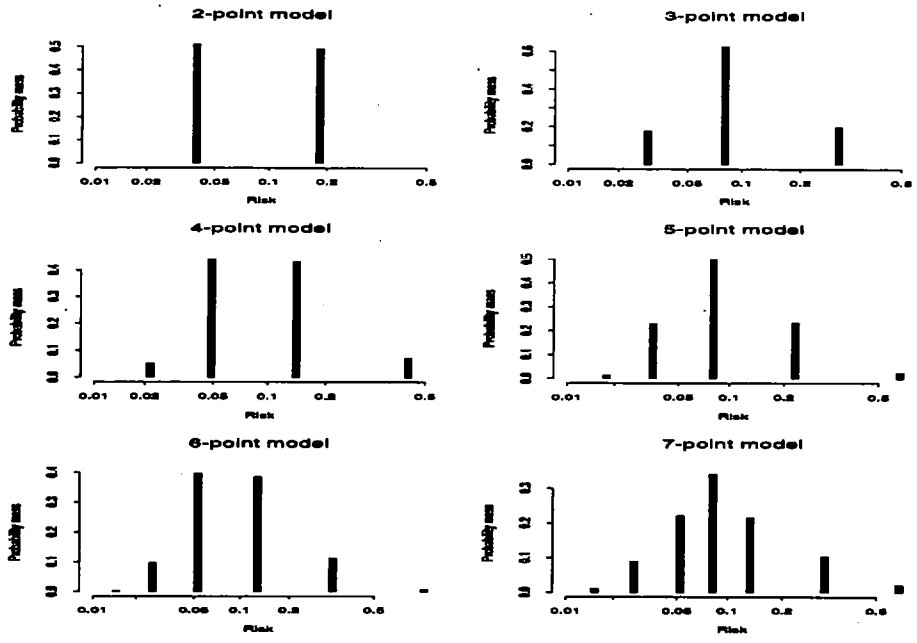


Figure 19 Non-parametric estimates of the distribution of hazard (on a logistic scale). Women. First attempt.

Number of masses	Deviance	Variance logit(λ)
1	6042.06	0.0
2	6016.31	0.735
3	6015.52	0.671
4	6015.09	0.737
5	6015.16	0.725
6	6015.18	0.735
7	6015.28	0.701

Table 42 Fitting non-parametric frailty distributions to the first attempt data

The convergence difficulties of the EM algorithm are apparent since the deviance does not decrease uniformly with the number of masses. Indeed, there is little to choose between any of the "solutions" with three or more masses. Nevertheless, this analysis gives a clear impression that the distribution of random effects is fairly symmetric on the logistic scale

with variance of around 0.7. There is no evidence for a proportion of sterile women, nor for a multimodal distribution of fecundability. This agrees quite closely with the variance estimates of 0.740 obtained from the beta-geometric model (likelihood estimate) and of 0.801 obtained from the logistic Gaussian model (Gibbs sampling estimation).

The same model is fitted to the male hierarchy, with introduction of cycle rank and a dummy variable for attempts (0 if first, 1 if subsequent). This introduction of cycle rank and attempt rank give us a way to model marginally (population-averaged) with respect to women, but conditionally (subject-specific) with respect to the donors. Before any introduction of the sperm characteristics, the distribution of random effects shows the existence of a percentage of rather infertile donors (Figure 20, first graph, on the left, about 10p100 of mass on a mass-point at about 0.01). But observed covariates are able to identify these donors as shown by the two other graphs , presenting respectively the estimated distribution of residual heterogeneity between donors after introduction of covariates (mobility, number and quality index) at donation level (about 5p100 of the women, at a probability of success being lower then 0.01) and after introduction of these covariates at donor level (no mass-point below 0.08 percent of success).

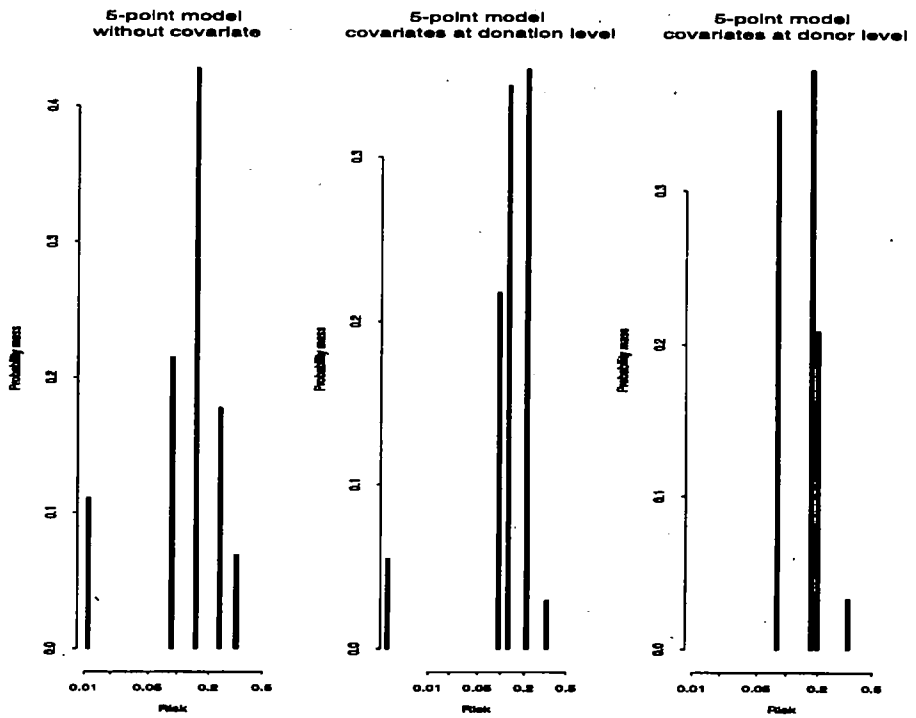


Figure 22 Non-parametric estimates of the distribution of hazard (on a logistic scale) Donors. Complet data set. Note that the fifth mass-point does not appear on the right hand graph the two lower mass-point being very close together at 0.081 and 0.0811

Table 43 shows the numeric values of the estimated mass-points and masses for a five point model including covariates at donor level.

Parameter	Estimate	(s.e.)	Mixture proportions
Attempts:			
Subsequent	0.265	(0.074)	
Cycle rank:			
2	-0.306	(0.103)	
3	-0.273	(0.107)	
4	-0.315	(0.114)	
5	-0.476	(0.125)	
6	-0.498	(0.134)	
7	-0.542	(0.148)	
8	-0.286	(0.143)	
9	-0.640	(0.174)	
10	-0.631	(0.188)	
11	-0.906	(0.216)	
12	-0.931	(0.237)	
Donor :			
Average Sperm count	0.201	(0.041)	
Average Sperm mobility	0.185	(0.035)	
Average Sperm quality	0.349	(0.046)	
Heterogeneity (Donors) :			
Mass points on risk scale			
First	0.0810		0.0258
Second	0.0811		0.3529
Third	0.1745		0.3793
Fourth	0.1979		0.2090
Fifth	0.3515		0.0330
Deviance	7534.4		

Table 43 Complete data. Non-parametric model for heterogeneity between donors

These results confirm those of the Chapter 7. The fixed effects of the characteristics of the donations explain an important part of the variance among donors : The mass of the lower mass-point of the distribution of residual heterogeneity decreases when the fixed effects are introduced in the model. Moreover when introducing the mean of these covariates over all donations of a donor in the model rather than the separate values of these characteristics for

each specific donation this mass-point disappears : sterile donors are identified by the covariates introduced in that way.

2. Empirical Bayes Estimates

In this Section we outline the question of the potential use of empirical Bayes estimates of the random effects in clinical practice. The prediction of breeding values has motivated numerous works on random effects models (e.g., Henderson, 1984; Gianola et al, 1986; Foulley et al, 1987; Foulley et al,1990). We can discuss the interest of similar approach for human.

Only donors with some evidence of fertility are accepted and they are not discarded after several failures since it is difficult to attribute the failure to the lack of fertility of their sperm rather than to the male fertility of the women for whom it was used. In addition insemination clinics have few donors available and drastic method of exclusion would be counter-productive. Nevertheless the broad range of the success rate per donor calls for a discussion (50 donors where systematically unsuccessful in all cycles, over 1 to 71 cycles). Some donors have so bad results that it would probably be better to discard them from the sperm bank for ethical and practical reasons. Knowing past bad results of some of them, the posterior probability of success in further trials is probably very low. Two propositions could be made. Either to discard donors having low estimates of posterior probability of success in regard to a threshold, or discard systematically a percentage of the less fertile donors. A parallel can be drawn between the comparisons between donors and the comparisons of 'Institutional performance' discussed recently by Goldstein and Spiegelhalter (1996). These authors point out the interest of 'adjusted comparison' of

organisations in Educational or Health systems, but also stressed the inevitable limitations in making such comparisons.

We apply the two methods proposed in their paper to the donors. First following Goldstein (1995b) we present on the right hand side of Figure 23 the empirical Bayes estimates of the linear component for the 279 donors, with confidence intervals, which enable pairwise comparisons to be carried out while maintaining an average required type I error rate. This is achieved using as covariance matrix an estimation of the conditional covariance matrix $E\left[\left(\hat{b}-b\right)\left(\hat{b}-b\right)\right]$ and writing the confidence interval for the *i*th donor as given by $\hat{b}_i \pm z_\beta \sigma_i$, where z_β is the average of $z_\alpha \frac{\sigma_{ij}}{\sigma_i + \sigma_j}$ taken over all the pairs (i,j). z_α is the normal deviate with a two tail probability, and σ_i, σ_j and σ_{ij} are respectively the standard deviation of both donors and of the difference between them.

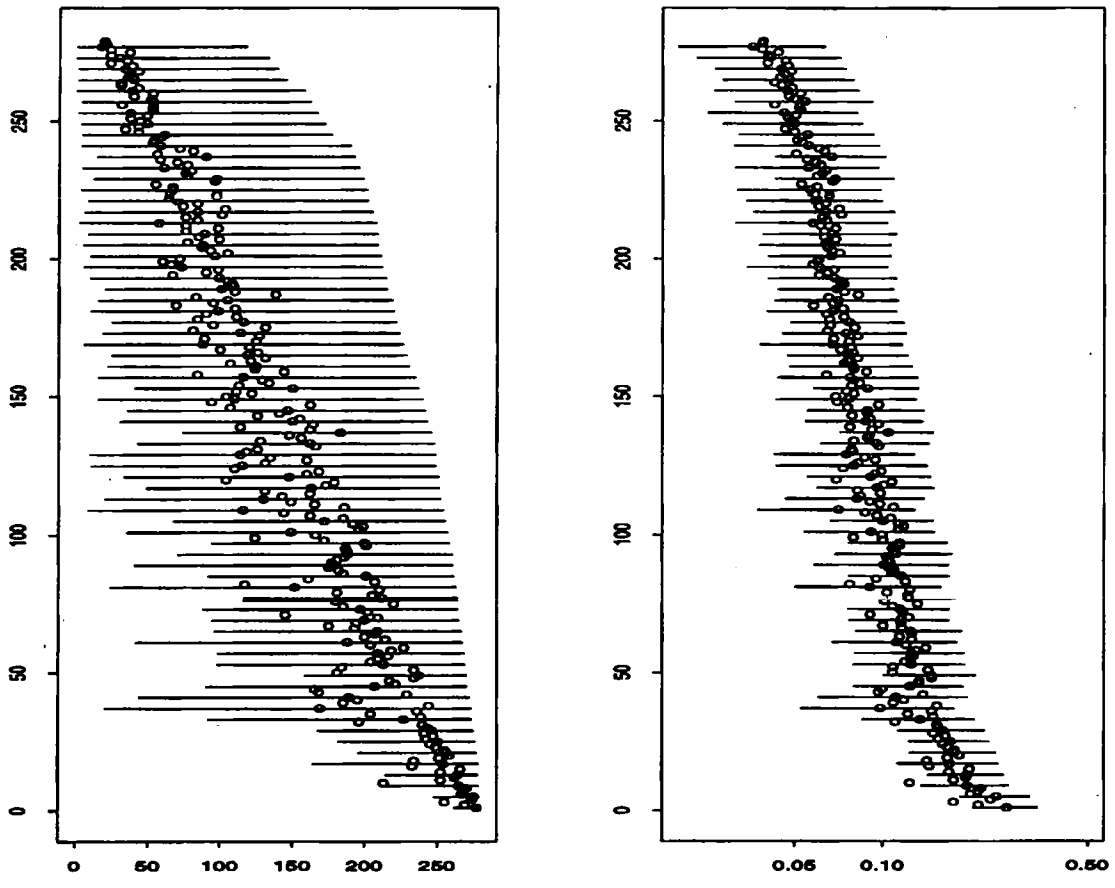


Figure 23. Empirical Bayes estimates of the success rate of the donors. Left : Median ranks and 95% confidence intervals. Right : pairwise 95% overlap intervals. Logistic mixed model with the characteristics of the donations as fixed effects and donors effects as random. In the two figures donors are sorted identically.

The rank of the estimated random effects are obtained fitting the logistic Gaussian mixed model using BUGS.

BUGS code is quite similar to that presented in the previous Chapter for logistic Gaussian mixed model plus the following lines (Spiegelhalter, 1995)

```
# Compute ranks :
for (j in 1:D) {
   for (k in 1:D) {greater.than[j,k] <-step(predict[k]-predict[j]);}
   rank[j]<-sum(greater.than[j,]); # rank of the donor j
}
```

The median estimates and 95% confidence interval are displayed on the left hand side of Figure 23. The width of the intervals is noteworthy. There are obvious relations between *MLn* estimates and BUGS ranks. The same is true for intervals.

The question of optimal selection of donors having to be discarded from the sperm bank need to be investigated more completely. An optimal solution would take into account the limitation of the number of born children per donor : for genetical reasons the French law limits to 5 the number of living children per donor. When a donor has being used for 5 successful inseminations with live-birth his semen must be discarded from the bank. Despite the difficulty to recruit donors it is necessary to discuss some selection, because it would not be ethically acceptable to use sperm known to be inefficient. This question of practical interest merits further investigation.

3. Open problems and discussion

In this Section we sketch some other open problems and conclude this dissertation stressing the interest of mixed models to analyse fecundity data.

Accuracy of approximations

In the previous Chapters we have emphasized the difficult choice we had to make between the estimates of variance components obtained using *PQL*, this method providing results in a few minutes despite the large size of the dataset, and the probably better estimates obtained using Gibbs sampling but precluding quite absolutely to use them in the building model phase, the computing time being really too long.

PQL was clearly for AID dataset the method of choice to build up the model, particularely thanks to the existence of a flexible software — *MLn* — allowing, for example, to test for

the existence of interactions between fixed effects and variance components. But MCMC methods were necessary to obtain corrected estimates, particularly of the variance components. *PQL* bias of the variance of random effects estimate is larger when the variance components are higher. Breslow and Lin (1995) have shown, using simulation technics for a single random effect, the existence of a asymptotic bias of $\approx$-0.4 for standard deviation of 1. In AID dataset we observe *PQL* estimated values of about 0.5 (Table 35) for the woman variance component, to compare with about 1 (Table 42) obtained using Gibbs sampling. The bias is smaller for donor variance component this variance being itself smaller than the woman's. A second reason of the difference of size of bias between female and male variance component could be the cluster size. Performance of *PQL* is less satisfactorily when the data are sparse (Breslow and Lin, 1995). In AID dataset the mean number of cycles per woman is 4.65 to compare with 31.65 for the donors. Moreover an added difficulty could be due to the censored nature of AID data. The number of success per woman and donor are respectively 0.48 and 3.26. The effect of censoring concerns mainly the woman. In AID dataset 52 % of the women did not conceive. For censored observations likelihoods for random effects are monotone (Clayton, in Discussion of Lee and Nelder paper, 1996) so that the Gausian approximation which underlies *PQL* will fail. A last reason for a bias can be the fact that women are removed from observation after conception. Considering *PQL* as a "Predicted" quasi-likelihood approach helps to explain this potential particular effect. *PQL* is a method of estimation based on a linearization of the link function around the predicted value. In the "delay to event" situation the cluster size depends heavily on the probability of success : higher is the probability of success, lower is the cluster size and thus more important is the shrinkage; empirical Bayes estimates may not be the better point of linearization in this case ? This last consideration needs more work to be confirmed.

Some authors have proposed solutions to improve approximations for multilevel models with binary responses. Breslow and Lin (1995) propose a method to correct the bias in generalized linear models with a single component of dispersion, but also for multiple components (Lin and Breslow 1997) using a correction terms. Their simulations tends to prove the interest of this approach. Kuk (1995) proposes a Monte Carlo method for iterative bias correction. Recently Steele (1996) has proposed a modified *EM* algorithm for estimation in generalized mixed models whose results seem promising.

Goldstein (1995) and Goldstein and Rasbash (1996) propose an other method based on the addition of a second order term in the linearization of the link function. But the justification in terms of approximation of the likelihood remain to be explored. This second order approximation is proposed as an option in *MLn*. In Table 44 we present the effect of this correction on the results obtained when fitting to our dataset a logistic Gaussian mixed model with woman random effect. The estimations of the fixed effects and variance components are both higher than those obtained using *PQL* (Table 32). The estimation of the variance component is close to the one obtain fitting the data using Gibbs sampling.

Parameter	First Order approximation Estimate (s.e.)	Second Order Estimate (s.e.)
Intercept	-2.237(0.039)	-2.460(0.046)
Woman :		
Age (woman)	-0.105(0.037)	-0.132(0.043)
Azoospermia (husband)	0.080(0.037)	0.089(0.044)
Cycle :		
Insler score	0.260(0.039)	0.285(0.044)
Early insemination	-0.137(0.038)	-0.149(0.043)
Late insemination	-0.084(0.033)	-0.087(0.037)
Clomiphene citrate	-0.104(0.036)	-0.108(0.041)
Donation :		
Sperm count	0.140(0.030)	0.153(0.033)
Sperm motility	0.175(0.033)	0.195(0.037)
Sperm quality	0.249(0.036)	0.267(0.040)
Heterogeneity :		
Between women	0.569(0.074)	0.979(0.105)

Table 44 Complete data. Female hierarchy

Table 44 shows the results of the same second order correction in the case of two random effects, one for the women and one for the donors. Again the results are modified. They are close to those obtained with Gibbs sampling (Table 35). As shown Table 45 this does not decrease the approximated log-likelihood because these second order estimates do not maximize the penalized quasi-likelihood. But it would be interesting to investigate this method further.

Parameter	First Order approximation estimate (s.e.)	Second Order Estimate (se)
Intercept	-2.303	-2.509(0.045)
Woman :		
Age (woman)	-0.106 (0.036)	-0.131(0.042)
Azoospermia (husband)	0.080 (0.037)	0.090(0.043)
Cycle :		
Insler score	0.264 (0.039)	0.287(0.044)
Early insemination	-0.137 (0.038)	-0.149(0.042)
Late insemination	-0.082 (0.034)	-0.085(0.037)
Cloniphene citrate	-0.103 (0.036)	-0.106(0.040)
Donation :		
Sperm count	0.130 (0.030)	0.142(0.033)
Sperm motility	0.179 (0.034)	0.197(0.037)
Sperm quality	0.217 (0.036)	0.231(0.039)
Heterogeneity :		
Between women	0.500 (0.072)	0.879(0.100)
Between donors	0.222 (0.043)	0.236(0.046)
-2 log likelihood (PQL approximate)	6 880.6	6986.8

Table 45 Complete data. Alternating EM algorithm. With covariates.

Finally, approximation methods and MCMC methods can probably be used both in a same estimation method : approximate estimates can be refined by sampling methods. This is a field for further research.

Analysis of data on human fertility

In this dissertation we have explored methological approaches adapted to correlated binary data and assessed their suitability for the analysis of data on human fertility. Studies of

intra-uterine insemination with donor's sperm presents some of the most challenging statistical aspects of fecundability studies:

[i] the delay until conception is subject to censoring since some women stop after a few cycles without success, choosing instead adoption or another treatment such as in vitro fertilization;

[ii] the heterogeneity of womens' fecundability is the source of a marginal decrease of the success rate from cycle to cycle, the most fertile women conceiving earlier;

[iii] after a first success, with the conception and the birth of a child, a second series of inseminations can be attempted: indeed several such programmes have been attempted by the same woman;

[iv] insemination with sperm obtain from donors introduces dependence between the outcomes of inseminations in which sperm from the same donor has been used.

Other clinical situations in human fertility are simpler : except for in vitro fertilization with sperm from donor female and male components of the fecundity are confounded. Unit specific regression models with closed form of marginal presented in Chapter 5 are certainly the solution of choice for the largest part of the current studies on human fertility. The gamma-geometric model with the Poisson approximation of the likelihood provides a very interesting solution for these analyses. The *rpoisson* macro — Personal communication, D.G. Clayton — in STATA may be proposed for users. Other members of the family of distributions described in Chapter 5 can be used when necessary. Indeed, Heckman and Singer (1984) showed substantial changes in parameter estimates with small changes in mixing distribution specification. It would be interesting to study further the interest of the inverse Gaussian family of distribution in the context of fecundability data. Chapter 5 showed also the particular interest of the complementary log-log link function. Nevertheless other link functions may be investigated. Animal breeders use mixed models

to separate fixed effects, for example age, breed or sex, from random genetic effects and other source of variation. Binomial traits are sometimes regarded as resulting from classifying an underlying normal variable into two classes relative to some threshold. Gianola and Foulley (1983), Gilmour Anderson and Rae (1985) and others have investigated these mixed models with probit link function. It could be useful to study the applicability of these methods to human fecundity.

It would also be worth to explore the non-parametric method of Aitkin (1996) for *crossed* random effects and the Yashin and Iachine (1995) "additive frailty components" approach.

More than randomized trials, observational data analysis can profit from mixed models. Observational data are remarkable in that a lot of covariates have to be taken into account and that a careful examination of interactions and of possibly random effect of the covariates have to be studied. This is a reason to stressed the interest of models including more than one random effect. Logistic Gaussian mixed models with more than one random effect fitted using a *PQL* algorithm such as those provided by *MLn* is of a greatest interest in this field. It allows to take benefit of the whole knowledge of the structure of the dataset without the limitation of applications limited to one random effect.

In our experience, contruction of compositional covariates describing the mean value or other characteristics such as the variability of covariates in repeated measurement (e.g. cycles) of a same unit (e.g. woman or donor) provide an important tool for investigation of the potential reasons of successes and failures of insemination. Such method of regression has probably a great interest in clinical epidemiology. Also the "compositional covariates" results raise the problem of covariate measurement error (Carroll, Ruppert and Stefanski, 1995). This methological problem has much in common with mixed effect models. Such models provide potential areas for future work.

Modeling of repeated measurement data such as AID data is not an enterprise that should be undertaken lightly, but is clearly of considerable interest.

REFERENCES

AALEN, O.O. (1978). Non parametric inference for a family of counting processes. Ann. Statis., 6,701 - 26.

AITKIN, M., ANDERSON, D., and HINDE, J. (1981). Statistical modeling of data on teaching styles (with discussion). J.R. Stat. Soc. A, 144, 148 - 61.

AITKIN, M., and FRANCIS, B.J.(1995). GLM4 macros for non parametric maximum likelihood estimation with overdispersion and variance component models in the exponential family. The GLIM Newsletter, 25, 37-45.

AITKIN, M. (1996). A general maximum likelihood analysis of overdispersion in generalized linear models. Statistics and Computing, 6, 251-62.

ANDERSEN, P.K., BORGAN, Ø., GILL, R.D. and KEIDING, N. (1992). Statistical models based on counting processes. Springer-Verlag. New York.

ANDERSON, D.A., and AITKIN, M. (1985). Variance component models with binary responses interviewer variability. J.R. Statist. Soc. B, 47, 203 - 10.

BAIRD, D.D., WILCOX, A.J., and WEINBERG, C.R. (1986). Use of time to pregnancy to study environmental exposures. American Journal of Epidemiology, 124, 470 - 480.

BEST, N.G., SPIEGELHALTER,D.J., THOMAS, A., and BRAYNE,C.E.G. Bayesian analysis of realistically complex models. J. R. Statist. Soc. A, 159, 323-42.

BOLDSEN, J.L., and SCHAUMBERG, I. (1990). Time to pregnancy. A model and its application. J. Biosoc. Sci, 22, 255 - 262.

BONGAARTS, J. (1975). A method for the estimation of fecundability. Demography, 12, 645-60.

BRESLOW, N.E. (1984). Extra-Poisson variation in log-linear models. Appl. Statis, 33, 38 - 44.

BRESLOW, N.E., and CLAYTON, D.G. (1993). Approximate inference in generalized linear models. J. Am. Statist. Ass. 88, 9 - 25.

BRESLOW, N., and LIN, X. (1995). Bias correction in generalized linear mixed models with a single component of dispersion. Biometrika, 82, 81 - 91.

BROOK, P.F., BRATT, L.R., and COOKE, I.D. (1994). The more accurate timing of insemination with regard to ovulation does not create a significant improvement in pregnancy rates in donor insemination program. Fertility and Sterility, 61, 308 - 313.

BRYK, A.S., and RAUDENBUSH, S.W. (1992). Hierarchical linear models : Applications and data analysis methods. Sage publications, Newbury Park.

CARROLL., R.J., and RUPPERT D. (1982) Robust estimation in heteroscedastic linear models. Annals of statistics, 10, 429 - 441.

CARROLL., R.J., RUPPERT D., and STEFANSKI, L.A. (1995). Measurement error in nonlinear models. Chapman and Hall, London.

CASELLA, G. and GEORGE, E.I. (1992). Explaining the Gibbs sampler. m. Statisi, 46, 167-74.

CECOS Fédération, and Lelannou, D. (1987). Bilan d'activité des CECOS. Contraception Fertilité Sexualité, 15, 693 - 5.

CHATFIELD, C., and GOODHART, G.J. (1970). The beta-binomial model for consumer purchasing behaviour. Appl. Statist, 19, 240 - 50.

CLAYTON, D.G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika, 65 ; 145 - 151.

CLAYTON, D.G. (1988). The analysis of event history data : a review of progress and outstanding problems. Stat. In Med., 7, 819-41.

CLAYTON, D.G. (1991). A Monte Carlo method for Bayesian inference in frailty models. Biometrics, 47, 467 - 85.

CLAYTON, D.G., and HILLS, M. (1993). Statistical models in epidemiology. Oxford University Press.

CLAYTON, D.G. (1994). Generalized linear mixed models. In W. Gilks, S. Richardson, and D. Spiegelhalter, editors, Markov Chain Monte Carlo in practice. Chapman and Hall, London.

CLAYTON, D.G. (1994). Some approaches to the analysis of recurrent event data, Stat Meth. Med. Res., 3, 244 - 62.

CLAYTON, D.G., and ECOCHARD, R. (1997). Artificial insemination by donor : discrete time survival data with crossed and nested random effects. In : First Seattle Symposium in Biostatistics : Survival Analysis, Springer Verlag Ed.

COLLINS, J.A. (1993). New treatments, preliminary results, and clinical practice. Fertility and Sterility, 60, 403 - 405.

CONOWAY, M.R. (1990). A random effects model for binary data. Biometrics, 46, 317 - 328.

COX, D.R. (1972). Regression models and life tables (with discussion). J.R. Statis. Soc. B, 34, 187 - 220.

COX, D.R., and HINKLEY, D.V. (1974). Theoretical Statistics. Chapman and Hall, London.

COX, D.R. (1984). Interaction - International Statistical Review, 52, 1 - 31.

CRAMER, D.W., WALKER, A.M., and SCHIFF, I. (1979). Statistical methods in evaluating the outcome on infertility therapy. Fertility and Sterility, 32, 80 - 86.

CROWDER, M.J. (1978). Beta-binomial anova for proportions. Appl. Statist., 27, 34-37.

CROUCHLEY, R. (1995) A random effects model for ordered categorical data. J.Am. Statist. Ass., 90, 489 - 98.

DAVID, G., and PRICE, W.S. (1980) (eds) Mode of Evaluation of results in Artificial insemination and Semen Preservation. Plenum Press, New-York.

DAVIDIAN, M., and GILTINAN, D.M. (1995). Non linear models for repeated measurement data. Chapman and Hall, London.

DEMPSTER, A.P., LAIRD, N.M., and RUBIN, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. J.R.Statist. Soc. B 39, 1 - 38.

DIGGLE, P., and KENWARD, M.G. (1994). Informative drop-out in longitudinal data analysis. Appl. Statist, 43, 49 - 93.

EDWARDS, A.W.F. (1972). Likelihood. Cambridge University Press.

Federation CECOS, SCHWARTZ, D., and MAYAUX, M.J. (1982). Result of artificial insemination in 2193 nulliparous women with azoospermic husbands. The New England Journal of Medicine, 306, 404 - 406.

FOOTE, R.H., and OLTEANU, E.A.B. (1980). Increasing fertility in artificial insemination by culling bulls or ejaculates within bulls. Proceeding 8th Tech. Conf. on A.I. and Reprod. NAAB, 6 - 11.

FOULLEY, J.L., IM, S., GIANOLA, D., and HOESCHELE, I. (1987). Empirical Bayes estimation of parameters for n polygenic binary traits. Genet. Sel. Evol., 19, 197-224.

FOULLEY, J.L., GIANOLA, D., SAN CRISTOBAL, M. and IM, S. (1990). A method for assessing extent and sources of heterogeneity of the residual variances in mixed models. J. Dairy Sci., 73, 1612-24.

GELFAND, A.E., and SMITH, A.F.M. (1990), Sampling based approaches to calculating marginal densities. J. Am. Statis. Ass., 85, 398.

GEMAN, S., and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. Pattn Anal. Mach. Intell.,6, 721-41.

GIANOLA,D and FOULLEY, J.L. (1983). Sire evaluation for ordered categorical data with a threshold model. Génétique, Sélection, Evolution 15, 201-24.

GIANOLA, D., FOULLEY, J.L. and FERNANDO, R.L. (1986). Prediction of breeding values when variances are not known. Genet. Sel. Evol., 18, 485-98.

GILKS, W.R., and WILD, P. (1992). Adaptative rejection sampling for Gibbs sampling. Appl. Statis., 41, 337 - 48.

GILKS, W.R., CLAYTON, D.G., SPIEGELHALTER, D. BEST, N.G., McNEIL, A.J., SHARPLES, L.D., and KIRBY, A. (1993) Modelling complexity : Applications of Gibbs sampling in medicine. J.R. Statist. Soc. B, 55, 39-52.

GILMOUR, A.R., ANDERSON, R.D., and RAE, A.L. (1985). The analysis of binomial data by a generalized linear mixed model. Biometrika, 72, 593 - 9.

GINI, C. (1924). Premières recherches sur la fecondabilité de la femme. Proceedings of the International Mathematical congres, 2, 889 - 892.

GOFFAUX, M. (1978). Intérêt de la sélection supplémentaire des taureaux ou des éjaculats en vue de l'amélioration du taux de non-retour moyen du centre d'insémination artificielle. Elev. et Insem. 165, 17 - 20.

GOLDSTEIN H. (1986) Multilevel mixed linear model analysis using iterative generalized least squares. Biometrika, 73, 43 - 56.

GOLDSTEIN, H. (1991). Non linear multilevel models with an application to discrete response data. Biometrika, 78, 45 - 51.

GOLDSTEIN, H., and RASBASH, J. (1992). Efficient computational procedures for the estimation of parameters in multilevel models based on iterative generalised least squares. Computational Statistics and Data Analysis, 13, 63 - 71.

GOLDSTEIN, H. (1995 a). Multilevel statistical models, 2nd edn, Kendall's Library of Statistics, London.

GOLDSTEIN, H., and HEALY, M.J.R. (1995 b). The graphical presentation of a collection of means. J.R. Statist. Soc. A 158, 175 - 177.

GOLDSTEIN, H. and RASBASH, J. (1996). Improved approximations for multilevel models with binary responses. J.R. Statist. Soc. A. 159, 385-443.

GOLDSTEIN, H., and SPIEGELHALTER, D.J. (1996). League tables and their limitations: Statistical issues in comparisons of institutional performance. J.R.Statist.Soc.A 159,Part3, 385-443.

GRAY R., LERIDON H., and SPIRA A. (1993). (eds) Biomedical and Demographic Determinants of Reproduction, Clarendon Press, Oxford.

GREEN, P.J., and SILVERMAN, B.W. (1993). Non parametric regression and generalized linear models : a roughness penalty approach. Monographs on Statistics and Applied Probability. London : Chapman and Hall.

GRIFFITHS, D.A. (1973). Maximum likelihood estimation for the beta-binomial distribution and an application to the household distribution of the total number of cases of a disease. Biometrics, 29, 637 - 48.

GUZICK, D.S., BROSS, D.S., and ROCK, J.A. (1982). A parametric method for comparing cumulative pregnancy curves following infertility therapy. Fertility and Sterility, 37, 503 - 507.

HARVILLE, D.A. (1974). Bayesian inference for variance components using only error contrasts. Biometrika, 61, 383-385.

HARVILLE, D.A. (1977). Maximum likelihood approaches to variance component estimation and to related problems. J. Am. Statist. Ass. 72, 320 - 340.

HECKMANN J.J., and WALKER J.R. (1990 a). Estimating fecundability from data on waiting times to first conception. J. Am Statist. Ass., 85, 283 - 295.

HECKMANN J.J., ROBB, R, and WALKER J.R. (1990 b).Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the method of moments. J. Am Statist. Ass., 85, 582 - 9.

HENDERSON, C.R., KEMPTHORNE, O. SEARLE S.R., and KROSIGK, C.M. (1959). The estimation of environmental and genetic trends from records subject to culling. Biometrics, 15, 192 - 218.

HENDERSON, C.R. (1984). Applications of linear models in animal breeding (University of Guelph, Guelph, Ontario, 1984).

HENRY, L. (1953). Fondements théoriques des mesures de la fécondabilité naturelle. International Statistical Review, 21, 135 - 51

HENRY, L. (1961). La fecondité naturelle : Observation - théorie - résultats. Population, 16, 625-36.

HINDE, J.P., and WOOD, A.T.A. (1990). Binomial variance component models with a non-parametric assumption concerning random effects. In R. Crouchley, editor. Longitudinal data analysis. Avebury, Aldershot, Hants.

HOUGAARD, P. (1984). Life table methods for heterogeneous populations : distributions describing the heterogeneity. Biometrika, 71, 75 - 83.

HOUGAARD, P. (1986a). Survival models for heterogeneous populations derived from stable distributions. Biometrika, 73, 387 - 96.

HOUGAARD, P. (1986b). A class of multivariate failure time distributions. Biometika, 73, 671 - 8.

HUBER, P.J. (1967). The behaviour of maximum likelihood estimates under non-standard conditions. In Proceedings of the Fifth Berkely Symposium on Mathematical Statistics and Probability, I, 221 - 223.

IRWIN, J.O. (1949). A note on the subdivision of χ^2 into componements . Biometrika, 36, 130.

JOHNSTON, R.C., KOVACS, G.T., LORDING, D. H., and BAKER, H.W. (1994). Correlation of semen variables and pregnancy rates for donor insemination : a 15 year retrospective. Fertility and Sterility, 61, 355 - 359.

KALBFLEISH, J.D., and PRENTICE, R.L. (1980). The statistical analysis of failure time data. J. Wiley & Sons, New York.

KUK, A.Y.C. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. J.R. Statist. Soc B., 57, 395 - 407.

LAIRD, N.M. (1978) Nonparametric maximum likelihood estimation of a mixing distribution. J. Am.Statist.Ass., 73, 805-813.

LAIRD, N.M., and WARE, J.H. (1982). Random effects models for longitudinal data. Biometrics, 38, 963 - 974.

LANCASTER, H.O. (1949). The derivation and partition of χ^2 in certain discrete distributions. Biometrika, 36, 117.

LAURITZEN, S.L., DAWID, A.P., LARSEN, B.N., and LEINER, H.G. (1990). Independence properties of directed Markov fields. Networks, 20, 491-505.

LANSAC, J. Artificial insemination with frozen donor semen : a model to appreciate human fecundity In : GRAY, R., LERIDON, H., and SPIRA, A. (1993) (eds.) Biomedical and demographic determinants of reproduction, Clarendon Press, Oxford.

LARSEN, U., and VAUPEL, J.W. (1993) . Hutterite fecundability by age and parity : strategies for frailty modeling of event histories. Demography, 30, 81 - 102.

LEE, Y., and NELDER, J.A.(1996). Hierarchical generalized linear models. J.R. Statist. Soc. B, 58, 616-56.

LERIDON, H. (1977). Human fertility : The basic components.Chicago: University of Chicago Press.

LERIDON, H., and SPIRA, A. (1984). Problems on measuring the effectiveness of infertility therapy. Fertility and Sterility, 41, 580 - 6.

LEXIS, W. (1879). Über die theorie der stabilität statistischer reichen. Jahrbücher für nationalökonomic und statistik, 32, 60 - 98.

LIANG, K.Y., and ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. Biometrika, 73, 13 - 22.

LIANG, K.Y., SELF, S.G., and CHANG, Y.C. (1993). Modelling marginal hazards in multivariate failure time data. J.R. Statist. Soc. B, 55, 441 - 53.

LIN, X. and BRESLOW, N.E. (1997) Bias correction in generalized linear mixed models with multiple components of dispersion. J. Am. Statist. Ass., 92, in the press.

LINDLEY, D.V., and SMITH, A.F.M. (1972). Bayes estimates for the linear model. J.R. Statist. Soc. B, 34, 1 - 41.

LINDSAY, B.G. (1983) The geometry of mixture likelihoods : a general theory. Annals of Statistics, 11, 86-94.

LINDSTROM, M.J., and BATES, D.M. (1988). Newton-Raphson and EM algorithms for linear mixed effects models for repeated measures data. J. Am.Statist.Ass.,83, 1014-22.

LITTLE,R.J.A., and RUBIN, D.B. (1987). Statistical analysis with missing data. New-York:Wiley.

LONGFORD, N.T. (1987) A fast scoring algorithm for maximum likelihood estimation in unbalanced mixed models with nested effects. Biometrika, 74, 812 - 27.

LONGFORD, N.T. (1988). A quasi-likelihood adaptation for variance component analysis. In Proc. Sect. Comp. Statist. Am. Statist. Assoc., 137 - 42.

LONGFORD, N.T. (1994). Logistic regression with random coefficients. J. Comput. Statist. Data Anal., 17, 1 - 15.

LORNAGE, J., MATHIEU, C. (1989). Automated videomicrographic analysis of the movement of human spermatozoa in seminal plasma : assessing the fertilizing ability of sperm. In : IVth International Congress of Andrology Monduzzi, Bologna.

MAJUMDAR, H., and SHEPS, C. (1970). Estimators of a type I geometric distribution from observations on conception times. Demography 7, 349 - 60.

MALLET, A. (1986). A maximum likelihood estimation method for random coefficient regression models. Biometrika, 73, 645-56.

MARUANI, P., and SCHWARTZ, D. (1983). Sterility and fecundability estimation. J. Theor. Biol., 105, 211 - 219.

MASON, W.M., WONG, G.M., and ENTWISTLE, B. (1983). Contextual analysis through the multilevel linear model. In S. Leinhardt (Ed.). Sociological methodology (pp. 72 - 103). San Francisco : Jossey-Bass.

MATHIEU C. (1993). Preparation du sperme in vitro et son application dans le cadre des procréations médicalement assistées. These Biologie Humaine Lyon 1.

McCULLAGH, P., and NELDER, J.A. (1989). Generalized Linear Models, 2nd edn, Chapman and Hall, London.

McCULLAGH, P. (1980). Regression models for ordinal data (with discussion). J.R. Statis. Soc. B 42, 109 - 42.

McGILCHRIST, C.A., and AISBETT, C.W. (1991). Regression with frailty in survival analysis. Biometrics, 47, 461 - 466.

McGILCHRIST, C.A. (1993). REML Estimation for survival models with frailty. Biometrics, 49, 221 - 225.

OAKES, D. (1982) A model for association in bivariate survival data. J. R. Statist. Soc. B, 44, 414 - 422.

OAKES, D. (1986) Semi-parametric inference in a model for association in bivariate survival time data. Biometrika, 73, 353 - 361.

OAKES, D. (1989) Bivariate survival models induced by frailtes. J. Am. Statist. Ass., 84, 487-93.

PATTERSON, H.D., and THOMPSON, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545 - 554.

PAUL, S.R., and PLACKETT, R.L. (1978). Inference sensitivity for Poisson mixtures. Biometika, 65, 591 - 602.

PICKLES, A. (1994). Generalizations and applications of frailty models for survival and event data. Stat. Meth. Med. Res, 3, 263 - 78.

POTTER, R.G.and PARKER, M.P. (1964). Predicting the time required to conceive. Population Studies, 18, 99 - 116.

RASBASH, J., PROSSER, R., and GOLDSTEIN, H. (1989). ML2 and LM3 : Software for Two Level and Three-Level Analysis. Users guide. London : Institute of Education.

RASBASH, J., and GOLDSTEIN, H. (1994). Efficient analysis of mixed hierarchical and crossed-classified random structures using a multilevel model. Journal of Educational and Behavioral Statistics, 19, 337 - 350.

RASBASH, J., YANG, M., WOODHOUSE, G., and GOLDSTEIN, H. (1995). Mln : Command reference guide. London : Institute of Education.

RAUDENBUSH, S.W., and BRYCK, A.S. (1986). A hierarchical model for studying school effects. Sociol. Educ., 59, 1 - 17.

RAUDENBUSH, S.W. (1994). A crossed random effects model for unbalanced data with applications in cross-sectionnal and longitudinal research. Journal of Educational Statistics, 18, 321 - 349.

RODRIGUEZ, G., and GOLDMAN, N. (1995). An assessment of estimation procedures for multilevel models with binary responses. J.R. Statis. Soc. A., 158, Part 1, 73 - 9.

RUBIN, D.B. (1976) Inference and missing data Biometrika, 63, 581 - 92.

S-PLUS User's Manual, Version 3.3 for Windows (1995). Seattle : Statistical Science, Inc.

SAS Software SAS Institute, SAS Campus Drive Cary, NC U.S.A.

SCHUMITZKY, A.(1991) Nonparametric EM Algorithms for estimating prior distributions. Applied mathematics and Computing, 45, 141-157.

SCHWARTZ, D., McDONALD, P.D.M., and HEUCHEL, V. (1980). Fecundability, coït frequency and the viability of ova. Population Studies 34, 397 - 400.

SEARLE, S.R., CASELLA, G., and McCULLOCH, C.E. (1992). Variance components. J. WILEY & Sons, New York.

SHALL, R. (1991). Estimation in generalized linear models with random effects. Biometrika 78, 719 - 27.

SHEPS, M.C. (1964) On the time required for conception. Population Studies, 18, 85 - 97.

SHEPS, M.C., and MENKEN, J.A. (1973) Mathematical models of conception and birth. University of Chicago Press, Chicago.

SINGH, S.N., and BHADURI, T. (1972). Maximum likelihood estimates for the parameters of a continuous time model for first conception. Demography, 9, 249 - 255.

SKELLAM, J.G. (1948) A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the set of trials. J.R. Statist. Soc. B, 10, 257 - 261.

SPIEGELHALTER, D.J., THOMAS, A., BEST, N.G., and WILKS, W. (1995). *BUGS: Bayesian inference Using Gibbs Sampling, Version 0.50* MRC Biostatistics Unit, Cambridge.

STEELE, B.M. (1996). A modified EM algorithm for estimation in generalized mixed models. Biometrics, 52, 1295-310.

TAN, S.L., ROYSTON, P., CAMPBELL, S., JACOBS, H.S., BETTS, J., MASON, B., and EDWARDS, R.G. (1992). Cumulative conception and livebirth rates after in-vitro fertilisation. Lancet, 339, 1 390 - 94.

TANNER, M.A. (1993). Tools for Statistical inference. New-York : Springer-Verlag.

TIETZE, C., GUTTMACHER, A.F., and RUBIN, S. (1950). Time required for conception in 1727 planned pregnancies. Fertility and Sterility, 1, 338 - 46.

TRUSSEL J. et al (1985). Determinants of birth-interval length in the Philippines, Malesia and Indonesia : a hazard model. Demography, 22, 145 - 69.

TRUSSEL, J., HANKINSON, R., and TILTON, J. (1992). Demographic applications of event history analysis. Oxford : Clarendon Press.

VAUPEL, J.W., MANTON, K.G., and STALLARD, E. (1979). The impact of heterogeneity in individual frailty and the dynamics of mortality. Demography, 16, 439 - 454.

WAKEFIELD, J.C., GELFAND, A.E., and SMITH, A.F.M. (1991). Efficient generation of random variates via the ratio of uniforms method. Statis. Comput, 1, 129 - 33.

WEDDERBURN, R.W.M.(1974). Quasi-likelihood functions, generalized linear models, and Gauss-Newton method. Biometrika, 61, 439-47.

WEI L, J., LIN, D.Y., and WEISSFELD, L. (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. J. Am. Statist. Ass., 84, 1065 - 73.

WEINBERG, C.R., and GLADEN, B.C. (1986). The beta-geometric distribution applied to comparative fecundability studies. Biometrics, 42, 547 - 560.

WEINBERG, C.R., WILCOX, A.J., and BAIRD, D.D. (1989). Reduced fecundability in women with prenatal exposure to cigarette smoking. American Journal of Epidemiology, 129, 1072 - 78.

WEINBERG, C.R., BAIRD, D.D., and ROWLAND, A.S. (1993). Pitfalls inherent in retrospective time-to-event studies : the example of time to pregnancy. Statistics in Medicine, 12, 867 - 879.

WEINBERG, C.R., BAIRD, D.D., and WILCOX, A.J. (1994). Sources of bias in studies of time to pregnancy. Statistics in Medecine, 13, 671 - 681.

WHITE, H. (1980) A heterskedasticity-consistent covariance matrix timate and a direct test for heteroskedasticity. Econometrika, 48, 817 - 830.

WILLIAMS, D.A. (1982). Extra-binomial variation in logistic linear models. Appl. Statist, 31, 144 - 8.

WOODS, J., and WEINSTEIN, M (1985). A model for age specific fecundability. Technical report, University of Michigan, Dep. of Anthropology.

YASHIN, A.I., and IACHINE, I.A. (1995). Genetic analysis of durations : correlated frailty model applied to survival of Danish twins. Genetic Epidemiology, 12, 529 - 538.

ZEGER, S.L., LIANG, K.Y., and ALBERT, P.S. (1988). Models for longitudinal data : a generalized estimating equation approach. Biometrics, 44, 1049 - 1060.

ZEGER, S.L., and KARIM, M.R. (1991). Generalized linear models with random effects ; a gibbs sampling approach. J. Am. Statist. Ass., 80, 79 - 86.

ZHOU, H., WEINBERG, C.R., WILCOX, A.J., and BAIRD, D.D. (1996). A random-effects model for cycle viability in fertility studies. J. Am. Statist. Ass., 91, 1413-22.